



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE HUMANIDADES
DEPARTAMENTO DE LETRAS VERNÁCULAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

ERICK MOURA RODRIGUES

GLOSSEMÁTICA NAS ENTRELINHAS: PROPOSTA DE TRADUÇÃO DE
***LANGUAGE* (1970), DE LOUIS TROLLE HJELMSLEV**

FORTALEZA

2024

ERICK MOURA RODRIGUES

GLOSSEMÁTICA NAS ENTRELINHAS: PROPOSTA DE TRADUÇÃO DE
LANGUAGE (1970), DE LOUIS TROLLE HJELMSLEV

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará como parte dos requisitos para obtenção do título de Mestre em Linguística. Área de concentração: Linguística.

Orientadora: Prof.^a Dr.^a Carolina
Lindenberg Lemos.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

R612g Rodrigues, Erick Moura.
Glossemática nas entrelinhas : proposta de tradução de Language (1970), de Louis Trolle Hjelmslev /
Erick Moura Rodrigues. – 2024.
321 f. : il.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-
Graduação em Linguística, Fortaleza, 2024.
Orientação: Profa. Dra. Carolina Lindenberg Lemos..

1. Hjelmslev. 2. Tradução. 3. Glossemática. 4. Semiótica. I. Título.

CDD 410

ERICK MOURA RODRIGUES

GLOSSEMÁTICA NAS ENTRELINHAS: PROPOSTA DE TRADUÇÃO DE
LANGUAGE (1970), DE LOUIS TROLLE HJELMSLEV

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará como parte dos requisitos para obtenção do título de Mestre em Linguística. Área de concentração: Linguística.

Aprovada em 17/04/2024

BANCA EXAMINADORA

Prof.^a Dr.^a Carolina Lindenberg Lemos (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Lopes Leite
Universidade Federal do Ceará (UFC)

Prof. Dr. Ivã Carlos Lopes
Universidade de São Paulo (USP)

À minha mãe, força motriz de tudo.

AGRADECIMENTOS

À minha orientadora, Prof.^a Dr.^a Carolina Lindenberg Lemos, pela disposição, acolhida e por ter aberto tantos novos caminhos neste percurso.

Aos professores Ivã Lopes e Ricardo Leite pelas caras contribuições durante as arguições da qualificação e da defesa.

Ao professor Ivanowik Dantas Valério pelo apoio essencial desde os idos de 2016 em nosso grupo de estudos.

Ao professor José Leite de Oliveira Júnior pelas dicas valiosas em meus estudos.

Aos meus amigos Denilson e Daniel, companheiros de longa data e interlocutores dos momentos de alegrias ímpares.

Aos meus amigos Elano, Paulo Henrique, Ericson, André, Eriston, Giuliana e Juliana, incentivadores, parceiros de alegrias, conversas profundas, planos e realizações.

Aos meus amigos Nathalie, Renan, Gleice, Laerton, Samiquele e Isaque pelas fugas da rotina e risadas sinceras.

Aos colegas do SEMIOCE, que tornam o aprendizado de semiótica uma experiência única.

À amizade de Leo Monteiro e nossos sonhos de infância.

À amizade de Fernando Gleibe e sua perspicácia inspiradora.

À Kauana Sales por nosso presente e por nosso futuro.

Ao meu irmão, Davi, parceiro e amigo de todas as conquistas.

À minha mãe, Luiza, pela educação e amor que me formaram.

A linguagem é o instrumento graças ao qual o homem modela seu pensamento, seus sentimentos, suas emoções, seus esforços, sua vontade e seus atos, o instrumento graças ao qual ele influencia e é influenciado, a base última e mais profunda da sociedade humana.

Louis Trolle Hjelmslev

RESUMO

Este trabalho tem como objetivo primário apresentar uma proposta de tradução da versão inglesa do livro *Language* (1970 [1943]), de Louis Trolle Hjelmslev, para o português do Brasil. Para tanto, elaboramos nosso processo de tradução ancorado na perspectiva metodológica de Berman (2013) para formular esquematicamente os processos de estruturação da tradução de *Language*. Como estamos lidando com a tradução da tradução, não devemos considerar o nosso processo tradutório como o ponto de partida, mas sim levar em conta o que já foi feito. Portanto, temos como base a tradução em inglês de Francis Whitfield além das traduções em francês de Michel Olsen (1966), espanhol de María Catalina (1968b) e em italiano de Anna Woolf (1970) como suporte. Ainda, sem o intuito de sermos exaustivos, incluímos trechos em dinamarquês (1973) da edição original em nosso estudo. Paralelamente a isso, nosso trabalho tem como hipótese que *Language* se constrói sobre uma glossemática que não se quer vista, ou, ao menos, não por seus contornos teóricos e metateóricos. Assim, é a partir da ausência do termo basilar – glossemática – constatada por Whitfield, bem como de nossa constatação da ausência, (apenas) aparente, das formas mínimas que a teoria estabelece em sua base explicativa, que formulamos também nossa proposta de tradução, que pretende, em segunda instância, analisar a construção dessa glossemática prevista para o público inexperto. A partir dos achados apreendidos e das reflexões proporcionadas pelo exercício de tradução, propomos ainda a expansão das fórmulas desenvolvidas por Hjelmslev ao incluir exemplos do português brasileiro, além de outras manifestações lusófonas, como o português falado em Angola. Dessa forma, ao propormos a tradução de uma obra que ainda não conhece os contornos de nossa língua, abrimos margem para a expansão dos estudos hjelmslevianos e seus desenvolvimentos. Em nosso estudo final, propomos analisar, ainda, a construção discursiva da definição de imanência conforme presente no livro *A linguagem em relação à construção empreendida nos Prolegômenos a uma teoria da linguagem* (2013), e observar como as escolhas argumentativas e operacionais revelam a manifestação ou o apagamento discursivo da imanência metodológica e epistemológica e como isso impacta a argumentação em cada livro.

Palavras-chave: Hjelmslev; tradução; glossemática; semiótica.

ABSTRACT

This work's primary objective is to present a translation proposal to the English version of the book *Language* (1970 [1943]), by Louis Trolle Hjelmslev, into Brazilian Portuguese. To this end, we developed our translation process anchored in the methodological perspective of Berman (2013) to schematically formulate the processes of structuring the translation of *Language*. As we are dealing with a translation of an already translated text, we should not consider our translation process as the starting point, but rather take into account what has already been done. Therefore, we rely on the English translation by Francis Whitfield in addition to the translation in French by Michel Olsen (1966), in Spanish by María Catalina (1968b) and the Italian translation by Anna Woolf (1970) as support. Furthermore, without the intention of being exhaustive, we included excerpts in Danish (1973) from the original edition in our study. Parallel to this, our work hypothesizes that *Language* is built on a glossematics that does not want to be seen, or, at least, not because of its theoretical and metatheoretical contours. Thus, it is based on the absence of this basic term – glossematics – noted by Whitfield, as well as our observation of the (only) apparent absence of the minimum forms that the theory establishes in its explanatory basis, that we also formulate our translation proposal, which intends, in second instance, to analyze the construction of this glossematics intended for the inexperienced public. Based on the findings learned and the reflections provided by the translation exercise, we also propose the expansion of the formulæ developed by Hjelmslev by including examples of Brazilian Portuguese, in addition to other Portuguese-speaking manifestations, such as the Portuguese spoken in Angola. Thereby, by proposing the translation of a work that does not yet know the contours of our language, we open space for the expansion of Hjelmslevian studies and their developments. Lastly, in our final study, we also propose to analyze the discursive construction of the definition of immanence as present in the book *Language* in relation to the construction undertaken in the *Prolegomena to a theory of language* (2013), and observe how the argumentative and operational choices reveal the manifestation or discursive erasure of methodological and epistemological immanence and how this impacts the argumentation in each book.

Keywords: Hjelmslev; translation; glossematics; semiotics.

LISTA DE TABELAS

Tabela 1 – Apresentação de uma “função” em <i>A linguagem</i>	156
Tabela 2 – Ocorrências da imanência como método nos <i>Prolegômenos</i>	163
Tabela 3 – Ocorrências da imanência epistemológica nos <i>PTL</i>	165
Tabela 4 – Definições operacionais em <i>A linguagem</i>	173

SUMÁRIO

1	INTRODUÇÃO	12
1.1	O lugar de <i>A linguagem</i> na obra de Hjelmslev.....	13
1.1.1	<i>A linguagem como uma obra de vulgarização: conexões com a linguística do século XIX</i>	18
1.2	O processo de tradução.....	21
1.3	Comentários e escolhas do processo tradutório.....	22
1.3.1	<i>Capítulos</i>	23
1.3.2	<i>Abreviações</i>	24
1.3.3	<i>Acréscimos, contribuições e substituições</i>	25
2	TRADUÇÃO – <i>A linguagem: uma introdução</i>	28
2.1	Introdução do tradutor.....	29
2.2	Observações preliminares.....	36
2.3	Função linguística.....	40
2.4	Relações genéticas entre as línguas.....	42
2.5	Estrutura linguística e o uso linguístico.....	62
2.6	Formação dos signos.....	73
2.7	Famílias linguísticas.....	93
2.8	Línguas originárias.....	104
2.9	Relação tipológica das línguas.....	111
2.10	Tipologia das estruturas linguísticas.....	116
2.11	Tipologia do uso linguístico.....	130
2.12	Mudança linguística.....	136
2.13	Diferentes graus linguísticos.....	145
2.14	Índice remissivo.....	150
3	ESTUDO CONTRASTIVO DA IMANÊNCIA NOS <i>PROLEGÔMENOS A UMA TEORIA DA LINGUAGEM E EM LANGUAGE</i>	159
3.1	Da imanência à <i>Imanência</i> : o desenvolvimento teórico nos <i>PTL</i> e em <i>Language</i>	162
3.1.1	<i>Function, fonction, funzione, función, funktion, função</i>	164
3.1.2	<i>A imanência nos PTL</i>	167

3.1.2.1	<i>A teoria da linguagem e a imanência.....</i>	168
3.1.3	<i>A imanência em A linguagem.....</i>	177
3.1.3.1	<i>Escolhas lexicais.....</i>	179
3.1.3.2	<i>Operacionalidade.....</i>	182
3.1.3.3	<i>Imanência.....</i>	185
4	CONCLUSÃO	188
	REFERÊNCIAS	194
	APÊNDICE A – ORIGINAL - <i>Language: An Introduction</i>.....	199

1 INTRODUÇÃO

A presente pesquisa terá como *corpus*, primariamente, o livro *Language* (1970). O referido livro faz parte do que se considera como o trabalho de divulgação ao público leigo tanto por seu caráter abrangente, como também pelo contexto de produção, cuja publicação assenta-se na chamada terceira fase de Hjelmslev (CIGANA, 2014). Nosso propósito primário será, então, a proposição de uma tradução que se organiza em torno da versão em inglês de *Language*. Essa proposta formula-se como uma expansão para o campo de estudos hjelmslevianos em território nacional que, em nossas letras, não possui ainda uma versão traduzida dessa obra bem como de muitas outras do teórico dinamarquês. De fato, em português brasileiro, encontramos apenas uma pequena parcela do arcabouço teórico de Hjelmslev. Temos a tradução de Teixeira Coelho dos *Prolegômenos a uma teoria da linguagem* (doravante, *PTL*), cuja primeira edição data de 1975, além da tradução do primeiro volume dos *Ensaio Linguísticos*, feita em 1991, por Antônio de Pádua Danesi. Com exceção desses dois títulos, tudo o que podemos acessar das obras de Hjelmslev está disponível apenas em outros idiomas. Assim, é a partir desse contexto que o presente trabalho se apresenta como uma proposição, e abertura, para que outras obras de Hjelmslev venham a ser traduzidas para nossa língua.

Além de abrir margem para novas traduções, o presente trabalho também vislumbra a expansão dos trabalhos produzidos acerca da teoria hjelmsleviana. Podemos já encontrar diversos textos que se debruçam sobre o arcabouço teórico de Hjelmslev como a tese de Carmo Jr. *Melodia & prosódia* (2007), que propõe um modelo para a interface música-fala a partir de um estudo comparativo entre o aparelho fonador humano e os instrumentos musicais reais e virtuais à luz da teoria glossemática de Hjelmslev. No campo da ampla semiótica, encontramos o estudo de Fiorin (2011) sobre a pretensa exclusão da história dentro dos estudos semióticos e como, a partir das proposições de Hjelmslev, pode-se ter uma perspectiva formal sobre a história. Em Bevidas (2015), encontramos uma primeira visada sobre a proposição de nível epistemológico que perpassa o texto dos *PTL*, perspectiva essa que foi posteriormente expandida no livro *Epistemologia discursiva* (2020), do mesmo autor. A tese *Condições semióticas da repetição*, de Lemos (2015), propõe uma revisão do papel desse conceito e como a perspectiva semiótica pode suprir as insuficiências identificadas quando esse conceito é operado a partir da retórica, por exemplo. Em pesquisa posterior, Lemos (2017) também investiga as condições formais necessárias para que o conceito da repetição possa ser

visto para além da semiótica puramente hjelmsleviana. Em estudos mais recentes, Valério (2023), em sua tese intitulada *A descrição glossemática de estruturas multissemióticas na chamada de capa*, propõe a descrição semiológica das chamadas de capa em jornais impressos a partir do contexto teórico e metodológico da glossemática. Lemos e Mafra (2023), na análise da canção enquanto faixa de um álbum, traz Hjelmslev à cena para estabelecer as bases descritivas entre classes e componentes. No capítulo “Estratégias enunciativas na produção do texto publicitário verbovisual”, publicado no livro *Linguagens na comunicação: desenvolvimentos da semiótica sincrética* (2009), Carmo Jr. procura soluções para a delimitação do conceito de semiótica sincrética, para isso utiliza em sua argumentação os conceitos de *sincretismo* e de *grau de intimidade*, o primeiro apresentado nos *PTL* (2013) e o segundo introduzido em *La categoría de los casos* (1978). O primeiro conceito apresenta-se como uma superposição entre dois outros conceitos sob uma condição de suspensão de diferenças, como no caso da suspensão das diferenças entre as grandezas /e/ e /i/, quando em posição átona final, em que se anulam as diferenças e prevalece o sincretismo (/dentI/ “dente”) (2009, p. 172). Para especificar o conceito anterior, Carmo Jr. apresenta o conceito de grau de intimidade, que se estabelece a partir dos conceitos de *coerência*, *incoerência*, *inerência* e *aderência*. Essas categorias serviriam, então, para graduar e especificar os níveis de sincretismo entre duas outras categorias. No mesmo esteio dos graus de intimidade, está a dissertação de Façanha (2024). O trabalho analisa as relações que se estabelecem entre as partes e sua totalidade, chamadas de relações mereológicas do livro *Os dragões não conhecem o paraíso* e utiliza como critério de análise os graus de intimidade de Hjelmslev.

A partir desse breve levantamento, vemos que a teoria glossemática conta ainda hoje com interesse no universo da linguística e semiótica. Acreditamos que a proposta da tradução de um texto clássico do autor dinamarquês compõe com esse interesse e contribui para que trabalhos, gerais ou específicos, sobre a teoria glossemática de Hjelmslev possam se tornar cada vez mais profícuos em nossa área.

1.1 O lugar de *A linguagem* na obra de Hjelmslev

Para situar nossa argumentação, devemos antes localizar a produção de Hjelmslev, que está vinculada diretamente ao Círculo Linguístico de Copenhague. Apresentaremos a seguir as obras que temos à nossa disposição em outras línguas, além do dinamarquês, a partir da organização elaborada por Cigana (2014) em três fases:

1. Os *Principes de grammaire Générale* (doravante, *PGG*) (1968a [1928]), marco da primeira fase.

2. De 1931 a 1939, temos a segunda fase formada pelas obras *Structure générale des corrélations linguistiques* (1933); *La catégorie des cas* (I e II) (1935); *Essai d'une théorie des morphèmes* (1936); *La structure des oppositions dans la langue* (1937), além dos ensaios mais recentes produzidos no período entre 1951 e 1959, no qual se inclui *Outline of the Danish Expression System with Special Reference to the Stød* (1973 [1951]); *Animé et inanimé, personnel et non-personnel* e *Sémantique structurale* (1956).

3. Na terceira fase temos os *Prolegômenos a uma teoria da linguagem* (2013 [1943]); além de *Language* (1970 [1943]) e o *Résumé d'une théorie du langage* (2014 [1975]), das quais a segunda obra será o objeto de nosso trabalho, enquanto a primeira servirá de apoio para comparações pontuais.

Em *Language* (1970 [1943]), desenvolvido no mesmo período criativo dos *PTL*, segundo a introdução de Whitfield, renomado tradutor para o inglês do linguista dinamarquês, Hjelmslev se debruça sobre as relações genéticas e tipológicas entre as línguas. Para isso, traz para a construção teórica as definições já aportadas em seus *PTL*. Dessa forma, a escolha de algumas formulações teóricas nos dois livros pode alimentar discussões sobre o intento de Hjelmslev em relação cada um deles e, em síntese final, à teoria da linguagem. Por exemplo, nos *PTL* há um direcionamento claro em relação à contração funcional de furtivos: um furtivo não *possui* uma função, mas a *contrai* (2013, p. 39). Esse direcionamento objetivo e que não quer deixar margem para ambiguidade lexical não tem a mesma força em *Language*, cuja elaboração teórica parece deslizar, propositadamente, em direção a um texto que vise um público menos acadêmico. Nesse sentido, o desenrolar teórico que nos *PTL* desenvolve-se metateoricamente, aplicando a si mesmo a hierarquia da teoria desde o princípio do empirismo, na elaboração das definições, aos retornos por catálise que aportam na imanência de caráter epistemológico. Em *Language*, o desenvolvimento do texto se dá de maneira mais implicativa e direta, o que desemboca nas escolhas lexicais mencionadas. Pode-se notar esse modo de proceder na descrição da *função-elemento* entre os elementos expressivos e como essa dependência regula a relação genética entre as línguas.

Neste ponto podemos então demonstrar que, entre cada elemento da expressão (EE) em uma língua e cada EE em outra língua geneticamente relacionada, existe uma dependência ou relação, ou, como acordamos chamá-la, uma

função. É, de fato, nessa função entre os EE nas diferentes línguas – essa função-elemento – que repousa a relação genética entre as línguas.¹

As escolhas lexicais recaem em termos que desembocam sempre dentro da própria teoria, ou seja, um proceder metodológico a partir de uma atitude centralizada na própria estrutura linguística – perspectiva que abordaremos mais diretamente no capítulo 5.

Também pode-se perceber esse mesmo estilo quando observamos as formulações em *Language* sobre os conceitos de arbitrariedade, adequação, além do princípio do empirismo, que, dotadas de rigor e concisão nos *PTL*, naquele ganham contornos mais concretos por meio dos exemplos (1970, p. 16).

Em *Language*, de início, a construção do princípio do empirismo elabora-se sobre uma base diádica: a teoria da estrutura linguística deve ser consistente e exaustiva em sua descrição (1970, p. 105). A simplicidade, conceito que possui nos *PTL* relevância sobreposta pelos dois outros conceitos, surge primeiro, em *Language*, junto à arbitrariedade e à adequação. Apesar dessa escolha operacional, o conceito possui o mesmo aporte significativo que temos nos *PTL*, porém sem a mesma classificação em níveis – ou etapas – que se sobrepõem na direção à descrição da linguagem. Dessa forma, Hjelmslev elabora sua teoria da linguagem sobre elementos mais concretos: os exemplos têm um tom extralinguístico – as fórmulas matemáticas por trás da construção de uma ponte – e estabelecem uma primeira aproximação entre uma realidade extra língua e outra, linguística, que a circunda e dela reflete características fundamentais.

Além disso, a própria conceituação não é fechada, ou seja, a consistência e a exaustividade da teoria da linguagem não compõem uma *classe* analisada, pelo menos não explicitamente como ocorre nos *PTL*. Dessa forma, adota-se um discurso mais ingênuo, provavelmente de caráter proposital para dar conta de um público com menos expertise na área. A formulação desse princípio não ocorre da mesma maneira nos *PTL*, como já apontamos. Aqui, Hjelmslev já nos apresenta um princípio: a descrição é feita sobre um conjunto de premissas exigidas pelo objeto analisado: tais premissas devem obedecer ao princípio do empirismo, ou seja, devem ser não contraditórias, exaustivas e tão simples quanto possível (2013, p. 11). Esse todo coeso é resultado de uma análise que,

¹ N. T.: Traduzido do original em inglês: “Now it can be shown that between each expression element (EE) in one language and each EE in another, genetically related language there is a fixed dependence or relationship, or, as we have agreed to call it, a function. It is, in fact, on this function between the EE’s of the different language, on this *element-function*, that genetic relationship between languages rests” (Hjelmslev, 1970, p. 13).

desde já, aplica a teoria a si mesma. Aqui temos uma classe analisada em outras classes e que formam, assim, uma hierarquia conceitual. Podemos entender o princípio do empirismo como uma classe submetida a uma análise, segundo essa perspectiva, a partir do momento em que a entendemos como pertencente ao conjunto teórico da glossemática. O empirismo, enquanto considerado como um objeto em contato com o extralinguístico, é um objeto não analisado, está fora da teoria. Segundo Hjelmslev, tal princípio não é regulado por sua etimologia ou percurso epistemológico, isso é explicitado quando ele nos diz que abandonará tal classificação caso a epistemologia o considere inadequado (2013, p. 11). Assim, podemos considerar que a arbitrariedade do princípio do empirismo dialoga com o próprio procedimento teórico da glossemática e exclui a possibilidade de que esse princípio seja estabelecido de outra maneira senão pela própria teoria.

A simplicidade, também, ganha um lugar bem definido na teoria, e de sua formulação terciária no princípio do empirismo, deriva seu próprio princípio: o da *simplicidade* (2013, p. 21).

Em decorrência da análise do princípio do empirismo, deriva da simplicidade, enquanto princípio, a única maneira pela qual podemos afirmar preferência por determinada ocorrência não contraditória e exaustiva em comparação com outra que também o é. Dessarte, a título de exemplificação, esse princípio dialoga com o conceito de simplicidade de Popper (2008) e, conseqüentemente, com o grau de falseabilidade de tais ocorrências ou soluções, ou seja, caso estejamos diante de duas soluções que possuem o mesmo grau de não contradição e de exaustividade, será preferida aquela que seja a mais simples. Tal simplicidade, no caso de Hjelmslev, diz respeito à extensão da análise. Se lidamos com as fórmulas para a concordância do elemento da expressão *m* em determinadas línguas indo-europeias, será preferível aquela de menor extensão, como ocorre no capítulo “Relações genéticas entre as línguas” de *A linguagem*. Portanto, a simplicidade se torna o princípio regulador do empirismo.

Além do mais, podemos elaborar uma breve passagem pela filosofia da ciência de Popper para apontar algumas semelhanças de desenvolvimento que podem nos ajudar a entender a construção do procedimento metodológico do princípio do empirismo de Hjelmslev. Em Popper, afastando-se do indutivismo – assim como Hjelmslev (2013, p. 13) –, a falseabilidade atua como um critério que define os limites entre o conhecimento científico e o conhecimento não-científico. No mesmo esteio de que as teorias nunca podem ser empiricamente verificáveis (Popper, 2008, p. 41), Hjelmslev, que nos diz que a atribuição do termo “empirismo” como nomenclatura para seu princípio é mera questão

de escolha (2013, p. 21), elabora sua teoria sobre uma base de verificação que compete somente à aplicabilidade dessa teoria, ou seja, aos dados resultantes da análise (ou os dados da experiência) não cabe a asserção da validade da teoria, mas somente sua aplicabilidade.. Assim, poderíamos supor uma proximidade entre os conceitos de falseabilidade e aplicabilidade. Badir (2014) já nos apresenta essa proximidade e comparação como historicamente possível e, ainda, apresenta outros pontos de concordância entre as perspectivas de ambos como a necessidade de uma base empírica e a distinção entre “universal em sentido estrito” e “universalidade numérica”, que se assemelham às distinções hjelmslevianas entre “universal” e “geral”. A construção desses princípios nos *PTL* torna-se, em suma, sucinta e marcada por uma parataxe de períodos curtos e objetivos.

Por outro lado, em *Language*, observamos que o andamento do texto torna-se menos propenso a picos de tonicidade² e adquire um caráter mais linear, o que consideramos ser mais uma forma de construir um texto academicamente mais abrangente. Como dito anteriormente, o princípio regulador de qualquer relação genética entre línguas manifesta-se como uma função geral que, no espectro de relações entre línguas, pode ser utilizada para definir outras funções que sejam mais específicas a cada caso, organizando-as hierarquicamente. Assim, uma única função que permeia todas as línguas do indo-europeu, por exemplo, é designada por um único símbolo: **m* (1970, p. 15). Tal símbolo servirá para designar a fórmula de um elemento-função que representa a concordância desse elemento entre as línguas indo-europeias. Nesse exemplo, observamos que essa função geral ocupa um nível hierárquico superior, visto que pode ser analisada em outras funções mais específicas, por exemplo as funções contraídas entre o *m* gótico e o *m* celta, representadas pela seguinte fórmula: gót. *m* φ celt. *m*. Nesse caso, Hjelmslev não define tais níveis relacionais como classes, estratos ou planos, mas sim como graus de parentesco. Mesmo assim, devemos observar que se trata, também, de níveis hierarquizados e organizados segundo uma lógica estrutural imanente (1970, p. 21).

Consequentemente, nosso trabalho tem como hipótese que o texto de *Language* propõe a construção de uma glossemática que não se quer vista, ou, ao menos, não por seus contornos teóricos e metateóricos. Assim, é a partir da ausência do termo

² Pensamos aqui no fluxo de apresentação dos conteúdos do texto como sujeitos a acelerações e desacelerações, picos acentuais e descidas átonas nos termos de Zilberberg (2011).

basilar – glossemática – constatada por Whitfield, bem como de nossa constatação da ausência (apenas) aparente, das formas mínimas que a teoria estabelece em sua base explicativa, que formulamos também nossa proposta de tradução, que pretende, em segunda instância, analisar a construção dessa glossemática prevista para o público inexperto.

1.1.1 A linguagem como uma obra de vulgarização: conexões com a linguística do século XIX

A linguística desenvolvida no século XX, até a Segunda Guerra Mundial, continuou a ser, essencialmente, histórica e traz consigo desenvolvimentos do que foi realizado no século anterior. Para Faraco (2005), o senso de sistema e as características do trabalho empírico realizado desde o manifesto de William Jones em 1786 marcam o início da concepção da linguística como uma ciência cujos traços perpassaram todos os séculos até serem sistematizados a partir do corte epistemológico performado pelo *Curso de Linguística Geral*, de Saussure, publicado em 1916, que carrega em si muitos traços do que foi feito no século anterior, traços estes que foram retrabalhados sob uma nova ótica: aquela que entende a língua a partir de uma consistência formal e sistematizada em uma totalidade organizada (Faraco, 2005, p. 28). Perceber o movimento engendrado por Saussure no *Curso* somente a partir da ótica de um rompimento – além de ser um ato precipitado –, deve ser feito com cautela, pois a construção do próprio livro excede o que Saussure, em si mesmo, entende por linguística. Isso é dito em função de a principal obra, em questão de alcance e divulgação, ser justamente cuja publicação não esteve sob os auspícios de Saussure, tanto por se tratar de obra póstuma, como por não ser de sua pena o texto final. Rastier (2017), em entrevista sobre Saussure e o desenvolvimento da semiótica à *Acta Semiótica et Linguística*, afirma que a dissonância e, portanto, a descontinuação de Saussure em relação ao que vinha se fazendo no século anterior, diz respeito à modificação da noção de objetivação, da construção do próprio objeto, que agora não se forma de maneira positiva nem ontológica, mas a partir da combinação de pontos de vista em dualidades – as dicotomias.

No prefácio à edição brasileira do *Curso de linguística geral* (2012), Salum destaca três pontos que ressaltam essa problemática: a primeira delas é o fato de que Saussure não estava satisfeito com o desenvolvimento da matéria (linguística), além de se sentir limitado pela própria compreensão de seus estudantes. Aliado a isso, ainda estava a sensação de que suas ideias não eram definitivas. O segundo ponto é que muito

difícilmente as anotações transformadas em livro corresponderiam ao que Saussure pensava e é analisada em “segunda mão”, pois acessamo-lo somente a partir de tais notas. Por fim, destaca a frase final do *Curso*: “a Linguística tem por único e verdadeiro objeto a língua encarada em si mesma e por si mesma”, cuja autoria não é de Saussure, mas sim de seus editores. Assim, a chave de ouro sistematizadora de Saussure não é uma formulação do teórico, mas um desenvolvimento de seus alunos, o que nos traz de volta ao apontamento inicial: apesar de nada disso ofuscar a importância e a radicalidade do movimento inovador de Saussure, é notável que o pensamento do linguista suíço e seu estruturalismo carregam consigo um teor mais historicista e de continuidade do que de rompimento como querem fazer acreditar algumas leituras de sua obra.

Um discurso de rompimento com essa tradição do século anterior vai se fazer presente a partir do momento pós-Segunda Guerra com Chomsky. Para Chomsky, o papel de uma teoria é validar uma gramática partir de um conjunto de dados não analisados, esse pressuposto, de que a teoria linguística é um procedimento de avaliação das possíveis gramáticas que geram as sentenças de uma língua, inaugura uma nova perspectiva para a linguística como ciência. Desse modo, o papel do linguista não é mais o de classificar os elementos de uma determinada língua com base em propriedades que podem ser observadas, mas sim o de compreender que tipo de procedimento *gerativo*, ou seja, que tipo de gramática pode gerar os fenômenos observados a partir dos dados empíricos selecionados e analisados. O rompimento engendrado pela Gramática Gerativa em relação ao Estruturalismo se deu a partir da ênfase no nível sintático. Se por um lado, para o Estruturalismo, as pesquisas deveriam focar nos níveis fonético, fonêmico e morfológico, por outro, para a Gramática Gerativa, a sintaxe ganha o status de mecanismo principal da gramática. É a sintaxe a responsável por duas propriedades universais da competência linguística humana: infinitude e criatividade. Entretanto, é incorreto descrever a Gramática Gerativa como uma ruptura total perante o Estruturalismo, como salienta Guimarães (2017, p. 84). O programa de pesquisa chomskyano manteve algumas das hipóteses do Estruturalismo. Por exemplo, é possível identificar na Gramática Gerativa a ideia de estrutura, formada a partir de um conjunto finito de objetos formais, e de restrições que regem a distribuição desses objetos na estrutura. A adoção dessas hipóteses é visível no modelo formal proposto em *Syntactic Structures* (1957). Essa concepção se alia à percepção de Altman (2021) em relação a um rompimento mais discursivo que teórico na construção da Gramática Gerativa de Chomsky, pois trazia em

seu estudo de 1957 características claramente estruturais: “uma concepção explícita, rigorosa e formalizada das proposições linguísticas” (Altman, 2021, p. 73).

Essa sistematização dos fatos trabalhados em linguística até o século XIX também se manifesta, segundo nossa concepção, dentro do escopo teórico de Hjelmslev, especialmente aquele que está situado nesse período pré-guerra. Nossa hipótese de base, baseada nas formulações que Whitfield apresenta em sua introdução ao livro, de que *A linguagem* possui um desenvolvimento teórico menos carregado – menos formalizante –, alinha-se a uma segunda hipótese: pode-se pensar que, a partir dessa construção discursiva mais operacional e, portanto, menos formalizante, teríamos um vislumbre mais nítido das influências e continuações do que fora feito no século XIX em matéria linguística.

Faraco (2005) elabora, no capítulo citado, um panorama dos principais pontos desenvolvidos desde o século XVIII a partir de William Jones e sua primeira visada sobre as semelhanças entre o sânscrito, o latim e o grego. Esse fora o marco inicial de uma sistematização progressiva da linguística, pois com o surgimento do método comparativo, surgia, conjuntamente, o pressuposto de que entre determinados elementos gramaticais de línguas aparentadas existiriam correspondências sistemáticas (Faraco, 2005, p. 32). Assim, paralelamente, podemos lembrar do caráter sistemático que Hjelmslev atribui às relações genéticas entre as línguas que são construídas sobre a base das relações existentes entre o que ele chama de *função-elemento*. Então, quando Hjelmslev afirma que “onde houver uma relação genética entre línguas, haverá função elementar entre os sistemas dessas línguas” (ver p. 51 *infra*), pode-se observar uma consonância com os estudos comparativistas. Em *A linguagem*, o estudo das mudanças e correspondências entre línguas carregavam um cerne imanente, por lidarem especificamente com mudanças linguísticas engendradas por elementos linguísticos, mas, em contrapartida, não se estendiam para além da pura comparação. É o que podemos observar no seguinte trecho, baseado nas leis de Grimm e Verner: “As consoantes /p/, /t/, /k/ do indo-europeu, quando precedidas de sílabas fracas, passam, no gótico, respectivamente a /b/, /d/, /g/; e, nos demais contextos a /f/, /θ/, /h/ respectivamente” (Faraco, 2005, p. 30). Em Hjelmslev, tais relações interlinguísticas baseiam-se em funções entre os chamados *elementos da expressão*. Assim, quando há uma função entre elementos da expressão de diferentes línguas, tem-se, portanto, uma relação genética, o que aponta para uma origem comum entre tais línguas e uma mesma *família linguística* (Hjelmslev, 1970, p. 9). Podemos observar, em suma, uma aproximação e expansão do que já fora feito. Outro ponto

interessante a ser observado nesse mesmo texto refere-se à concepção da linguística como ciência autônoma, ideia que está presente desde Whitney, que vislumbrava nesse intento uma diferenciação da linguística enquanto ciência do estudo histórico-comparativo. Esse é o mote tanto dos *PTL* como de *A linguagem*: a criação de uma linguística *linguística*, interessada pelo estudo da linguagem como um fim e não um meio que justifique outras concepções físicas, fisiológicas ou psicológicas (Hjelmslev, 2013, p. 2). Em suma, é importante entendermos Hjelmslev não somente a partir da ótica do rompimento de uma tradição, mas também a partir da atualização dessa tradição e *A linguagem* pode ser visto como esse livro que carrega mais nitidamente a tradição do que fora feito não só do século XIX mas também nos séculos precedentes.

1.2 O processo de tradução

Com base então nessa ideia de que *A linguagem* visa em sua estrutura um público mais amplo, elaboramos nosso processo de tradução ancorado na perspectiva metodológica de Berman (2013) para formular esquematicamente os processos de estruturação da tradução de *Language* (1970). Para o teórico da tradução, o ponto de partida da tradução é considerá-la como um constante devir que pressuporá uma *retradução*. A primeira tradução é naturalizadora, na medida em que introduz a obra estrangeira à cultura de chegada reduzindo a alteridade a fim de melhor integrá-la a essa outra organização cultural, afeiçoando-a a partir de estruturas socioculturais que buscam alcançar o destinatário. A retradução, por outro lado, faz um movimento de retorno ao texto original, no qual busca-se a afirmação cultural do outro na tradução, a própria singularidade do outro ganha relevância como ferramenta de fidedignidade da tradução. A tradução naturalizadora dá lugar, assim, a uma tradução que põe em relevo as especificidades linguísticas, estilísticas, textuais do texto-fonte, retraduzindo-o a partir das particularidades de sua estrutura cultural. Em nosso caso, como estamos lidando com a tradução da tradução, não devemos considerar o nosso processo tradutório como o ponto de partida, mas sim levar em conta o que já fora feito. Portanto, temos como base, fundamentalmente, a tradução em inglês de Francis Whitfield e, secundariamente, por questões que envolvem contrastes vocabulares, os quais aprofundaremos mais adiante, a tradução em francês de Michel Olsen (1966), em espanhol de María Victoria Catalina (1968b) e em italiano de Anna Woolf (1970).

À vista disso, lidaremos com descrição e tradução a partir de uma base não etnocêntrica. Porém, focaremos em algumas adaptações que visem desenvolver um texto aclimatado ao leitor falante de português, a partir da inserção sistemática da língua portuguesa nas funções e exemplos que forem convenientes ao nosso objetivo.

Em relação à nossa escolha acerca de qual tradução partiríamos para a elaboração da versão brasileira, cabe refletirmos sobre a importância de Whitfield – e de sua relação com Hjelmslev – na tradução da versão em inglês dos *PTL* e de *Language*. Em um curto texto publicado em 1966 na revista *Language*, Whitfield nos oferece um breve panorama biográfico de Hjelmslev. Após a publicação de um artigo escrito por Martinet (1946), houve um aumento acentuado na demanda por traduções do *Omkring sprogteoriens grundlæggelse* (1993b [1943]), até então apenas disponível em dinamarquês. E com essa demanda acentuava-se a preocupação de Hjelmslev sobre a recepção do livro e se a função proposta em sua concepção seria entendida adequadamente: a de ser um trabalho de popularização da teoria glossemática. Nesse aspecto, Whitfield salienta a proximidade de Hjelmslev com o trabalho de tradução da obra e o quanto o mestre dinamarquês dedicou-se, de forma generosa e cortês, à elaboração da versão final em inglês, cujo título ele próprio indicou. Entretanto, a tradução de *Sproget: en introduktion – Language*, em inglês – não pôde contar com a mesma felicidade, pois estava em processo no mesmo ano desta breve biografia de 1966, um ano após a morte de Hjelmslev.

Assim, nossa escolha pela versão inglesa apoia-se no fato de não termos acesso à língua original do texto, mas também no respaldo teórico e no diálogo que se estabeleceu entre Whitfield e Hjelmslev. Como apontado acima, nosso trabalho contará também com a tradução de Olsen para o francês a fim de resolvermos distinções lexicais que não estão presentes em línguas germânicas, como é o caso da oposição *língua vs. fala*, que não possui distinção em inglês, por exemplo. Nos casos em que encontrarmos tais dificuldades, levaremos em conta a versão francesa pela proximidade com o português e, acessoriamente, as versões espanhola de María Victoria Catalina e italiana de Anna Woolf pela mesma razão. Restando dúvidas, o contraste de escolhas entre duas traduções latinas pode mostrar caminhos mais pertinentes.

1.3 Comentários e escolhas do processo tradutório

Em relação às escolhas feitas, alguns aspectos importantes sobre a nossa proposta de tradução devem ser salientados. Em primeiro lugar, deve-se deixar claro que, pelo fato de usarmos a tradução francesa, e secundariamente a espanhola e a italiana, como apoio em nossa tradução teremos muitos ajustes que passarão pelo crivo da versão de Olsen, Catalina e Woolf Tais ajustes existem em virtude da proximidade que as línguas portuguesa, francesa, espanhola e italiana possuem, tornando a nossa versão mais adequada ao nosso propósito. Assim, incluiremos, neste tópico, nossas propostas de tradução que devem ser aprofundadas, seja em relação ao inglês, seja em relação ao francês, espanhol e italiano, além de escolhas que merecem pormenorizações.

1.3.1 Capítulos

Em relação à tradução dos títulos de capítulos, consideramos os seguintes procedimentos. No capítulo “Sign Formation”, encontramos em francês: “La formation des signes”. Esse plural é interessante, porque, em inglês, trata-se de uma construção adjetivada, em que não se usa plural (“Signs formation”, portanto, não funciona em inglês). Assim, “Formação dos signos” nos parece uma solução adequada. Já em português e em inglês, não precisamos de artigo em títulos, como acontece em francês, em que ele é sempre necessário. Para o capítulo “Parent Languages”, consultamos a versão francesa, na qual encontramos “Langues originelles” e decidimos adotar a formulação “Línguas originárias”. Já para o capítulo “Languages of Different Degrees”, em francês, encontramos a solução “Degrés linguistiques”. Uma escolha interessante, pois evita-se a questão língua ou linguagem. Assim, temos em português “Diferentes graus linguísticos”. A partir dessa padronização, seguimos o padrão estabelecido na versão em inglês para o restante dos capítulos. Dessa forma, temos “Função linguística” para “Linguistic Function”; “Relações genéticas entre as línguas” para “Genetic Relationship of Languages”; “Estrutura linguística e uso linguístico” para “Linguistic Structure and Linguistic Usage” e assim por diante. Por fim, escolhemos manter apenas a primeira letra dos títulos em maiúsculo, seguindo a padronização da ABNT, diferentemente da tradução de Whitfield.

Ainda sobre escolhas de tradução, preferimos utilizar “família linguística” para “Language Family” ao invés de uma construção locutiva. Também de igual importância é a escolha que fizemos pela tradução de *element-function*, que, em um primeiro momento fora traduzida como “função elementar”, cujo resultado decidiu-se

como não satisfatório, o que nos levou à versão francesa, na qual encontramos uma função de elementos (*fonction des éléments*), da qual se derivou a nossa proposta, função-elemento, preservando a solução com hífen do inglês, porém na ordem sintática do português.

No segundo capítulo, acreditamos ser pertinente destacar algumas passagens que poderiam constituir uma certa anacronia para o leitor atual. Desse modo, optamos por explicitar, por meio de notas de rodapé, informações que atualmente divergem do que fora apresentado na edição em inglês, como acontece em relação à população falante de línguas derivadas do indo-europeu. Outro ponto que merece destaque é o que apresenta o gótico como uma língua geneticamente aproximada, nesse caso, do dinamarquês, língua original da obra. Não sendo, portanto, próxima do português, decidimos especificar esse fato por meio de uma nota de rodapé.

No capítulo “Relações genéticas entre as línguas”, decidimos também adotar a exemplificação do material léxico (p. 38) a partir do latim, visto que o português não provê os elementos fonético-morfológicos para o argumento. Cremos, portanto, que nesse caso seja melhor adotar a solução do francês: recorrer ao latim. É mais próximo do português e resolve a questão. Nesse mesmo capítulo, optamos ainda pelas formas “romanzo” e “protoromanço” para designar as formas inglesas *romance* e *proto-romance*, respectivamente; conforme descreve Câmara (2011) em seu *Dicionário de Filologia e Gramática*.³ Por fim, utilizamos a *Gramática Grega* (2011), de Ragon, para solucionar dúvidas de tradução em relação às formas verbais gregas que surgem no capítulo “Relações genéticas entre as línguas”.

1.3.2 Abreviações

Em relação às abreviações, decidimos seguir o padrão das traduções latinas (francês, espanhol e italiano). Assim, todas as abreviações para os idiomas ficam em letras minúsculas. Registramos em maiúsculas também a abreviação para *elementos da expressão*. Para o último caso, salienta-se também que o uso da abreviação em letras maiúsculas torna mais simples a leitura das fórmulas apresentadas, pois esses fúntivos,

³ Em sua primeira edição, de 1956, o dicionário intitulava-se *Dicionário de fatos gramaticais*; já em 1963, na segunda edição, o título é alterado para *Dicionário de filologia e gramática*; e, por fim, em 1977, a obra assume a forma atual: *Dicionário de linguística e gramática*, ao qual fazemos referência neste trabalho.

para usar a terminologia dos *PTL*, formam a base de representação das fórmulas apresentadas no livro.

1.3.3 Acréscimos, contribuições e substituições

No capítulo “Relações genéticas entre as línguas”, Hjelmslev (1970, p. 22) apresenta a subfamília itálica pertencente ao tronco linguístico indo-europeu. Segundo essa classificação, incluem-se aqui as línguas românicas modernas. Na apresentação de uma função-elemento entre essas línguas, entretanto, o português não surge como exemplo no paradigma apresentado. Em uma primeira observação, pode-se considerar que essa escolha fora feita pois o elemento da expressão em análise tanto do francês – que aparece na fórmula – como o do português, possuem a mesma correspondência. Dessa forma, o português pode ter sido descartado como prevenção a uma redundância. Entretanto, como nosso objetivo primário é o leitor brasileiro, consideramos necessária a inclusão dos exemplos em português. No capítulo “Função linguística”, encontramos algumas ocorrências que exigiram comparações pontuais entre francês e inglês para se chegar a um resultado satisfatório. Assim, por exemplo, escolhemos, seguindo a versão francesa, a tradução “grupo de línguas” para a expressão em inglês *class of languages*.

Além disso, é necessário também apontar determinadas escolhas substitutivas que empreendemos em nosso percurso. Um pouco à frente, no capítulo “A estrutura linguística e o uso linguístico”, deparamo-nos com a exemplificação de determinados signos a partir de uma substituição paradigmática (p. 53). Para que a demonstração tomasse forma em português, decidimos pela inclusão de vogais ao final de cada exemplo para que uma segunda sílaba fosse formada, assim aproximando a construção dos exemplos da língua portuguesa. Assim, o que era *sat, rat, tat*, tornou-se *sato, rato, tato*. Ainda no mesmo capítulo, substituímos os exemplos *mig* e *jeg* – exemplos de signos não mais usados na língua inglesa – por *delo* e *quedo*, exemplos também de desuso, mas em nossa língua (p. 54). No subtópico “Formação proporcional”, substituímos os exemplos verbais do inglês pelo pronome indefinido *nada* e sua mudança de paradigma por substantivação (p. 72). Outros exemplos de substituição surgem na página 73 em relação ao processo de abreviação – substituído pelo exemplo *vossa mercê: você; motocicleta: moto*; UTI: *unidade de terapia intensiva*, dentre outros exemplos da mesma natureza. Em outros casos, optamos pela manutenção do exemplo em inglês (p. 75), seguindo o mesmo padrão apresentado na versão francesa. O português também foi usado nos trechos em

que se apresentam os elementos flexionais e os elementos radicos (p. 74). Cada um dos exemplos utilizados procura seguir o mesmo padrao do que encontramos nas versoes latinas de Hjelmslev.

Incluimos aqui tambem substituicoes pontuais em que nos foi permitido utilizar o portugues como contraponto ou complemento, como ocorre em relacao ao finlandes (p. 58), em que as palavras *skvat* e *skvulp* exemplificam as dificuldades que um falante lusofono teria para pronuncia-las e, em relacao ao ingles (p. 122), o par vozeado/desvozeado das palavras *zeal e seal* e, em portugues, *casa* e *caca*. Alguns comentarios tambem foram adicionados sem que os originais fossem excluidos. E o caso do breve comentario sobre o processo de gramaticalizacao do verbo *haver* (p. 74) e da indicacao dos modos verbais do portugues (p. 119).

Como ja ressaltamos acima, utilizamo-nos tambem das versoes em linguas romanicas de *A linguagem* como suporte para nosso estudo tradutorio. Assim ocorre quando traduzimos *feeling for a language* como *familiaridade lingustica* (p. 63), derivado do frances *familiarite avec une langue*. O mesmo ocorre quando traduzimos o subtopico “influencia dos elementos” (p. 64). Derivados do frances tambem sao as passagens em que encontramos as categorias *englobantes* (p. 115) e os elementos governados (*pressupostos*) (p. 117).

Alem do processo de traducao, tivemos que lidar tambem com a atualizacao de eventos e locais que, no caminhar da historia, foram alterados. Isso ocorre na traducao do subtopico da familia uralica do capitulo “familias lingusticas” (p. 90). Aqui, encontramos diversas referencias a Iugoslavia e a antiga Uniao Sovietica e sua Republica Socialista Sovietica Autonoma de Komi, hoje somente Republica de Komi. O mesmo ocorre no caso da antiga cidade de Gorky, atualmente conhecida pelo nome de Nijni Novgorod. Todos os casos em que houve qualquer mudanca de nomenclatura foram atualizados segundo a norma atual.

Sobre os achados apreendidos das reflexoes proporcionadas pelo exercicio de traducao, propomos analisar, ainda, a construcao teorica das definicoes presentes no livro em relacao a construcao empreendida nos Prolegomenos a uma teoria da linguagem (2013) e como essas escolhas impactaram na construcao argumentativa e operacional dos termos teoricos que surgem em *Language* e como elas revelam a presenca ou ausencia da

imanência metodológica e epistemológica. Dado que um tal estudo se apoia eminentemente na leitura de *A linguagem*, guardamos esse estudo para as últimas páginas desta dissertação.

2 TRADUÇÃO - A linguagem: uma introdução

Sumário

Introdução do tradutor

Observações preliminares

A função linguística

Relações genéticas das línguas

Estrutura linguística e uso linguístico

Formação dos signos

Famílias linguísticas

Línguas originárias

Relação tipológica das línguas

A tipologia das estruturas linguísticas

A tipologia do uso linguístico

A mudança linguística

Diferentes graus linguísticos

Índice remissivo

Introdução do tradutor

Embora o texto original em dinamarquês (intitulado *Sproget: En introduktion*) não tenha sido publicado até 3 de dezembro de 1963, o livro foi, na verdade, escrito vinte anos antes. Dessarte, aloca-se naquele período extraordinariamente criativo da vida de Hjelmslev – resplandecente com sua visão de uma nova “linguística linguística” – que produziu também os *Prolegômenos a uma Teoria da Linguagem* (*Omkring sprogteoriens grundlæggelse*) e, entre outros trabalhos que aguardam publicação póstuma, o *Résumé a uma Teoria da Linguagem* (*Sprogteori: Résumé*). Contrastando com a exposição compendiada e formal da “glossemática” elaborada no *Résumé*, ambos, *Prolegômenos* e *A Linguagem*, foram concebidos como popularizações: aquele, como um esboço dos fundamentos da glossemática, direcionado a uma audiência acadêmica; e este (no qual a palavra *glossemática* não surge uma única vez), como uma introdução aos problemas da linguagem e o seu estudo científico tendo em vista, primeiramente, o público leigo.

Há, de imediato, um notável contraste entre os dois trabalhos “de divulgação” que recaí no direcionamento a públicos distintos. Uma grande parte de *A Linguagem* concentra-se sobre a relação genética entre línguas, e a estrutura de todo o livro é determinada pela importância central que é dada às diferentes relações possíveis, genéticas e não genéticas, que as línguas podem ter umas com as outras. Nos *PTL*, por outro lado, a atenção é essencialmente direcionada às relações internas à língua e aos seus textos. Se, por outro lado, o leitor é levado a ver os dois livros, de alguma maneira, como demarcando os limites da linguística comparativa e da linguística descritiva, corre-se sério risco de se equivocar na leitura de ambos. Por um lado, as diversas discussões sobre as relações internas que o leitor irá encontrar no presente trabalho devem ser vistas como partes necessárias e orgânicas de seu argumento. Por outro lado, seria uma leitura superficial (ainda que comum) dos *PTL* que ignora o capítulo final e seu aviso expresso de que “não é a língua, em sua individualidade, o objeto do linguista.”

Para Hjelmslev, era óbvio que a linguística deve ser “comparativa”, não por puro comparatismo, mas para a busca exitosa de seus verdadeiros objetivos. A conclusão dos *PTL* deixa isso bastante claro, embora o corpo desse trabalho seja devotado – porque assim o podia, em vista de seu público particular – a outros aspectos da ciência linguística que seu autor almejava. Era igualmente evidente para Hjelmslev que uma introdução apropriada ao estudo da linguagem, tanto para o iniciante como para o linguista que

pretende revisar as bases de sua disciplina, deve passar pelo fio da gramática comparativa – particularmente a gramática comparativa do indo-europeu. Em sua resenha do trabalho *Foundations of Language*, de Louis H. Gray, Hjelmslev escreveu inequivocadamente:

Uma qualidade que, sem dúvida, será recebida com particular satisfação é que são levados em conta, em grande medida, fatos e teorias indo-europeus. O indo-europeu continua sendo o arcabouço do qual o linguista de várias épocas extrairá preferencialmente seus materiais e seus exemplos; além disso, o indo-europeu é o único domínio linguístico que *acreditamos* [itálicos adicionados por FJW] realmente conhecer a fundo. Portanto, é inútil querer poupar o leitor das dificuldades técnicas da gramática comparada; isso seria enganá-lo e privá-lo do essencial; pelo contrário, devemos encarar o problema de frente e fazer o que fez Gray: apresentar os problemas da linguística geral por meio dos fatos indo-europeus. Somente esse processo permitirá ao leitor compreender que, apesar de tudo, a linguística é uma ciência, se não na teoria, pelo menos na técnica, e penetrar, como observador, no laboratório do cientista. Nada é mais perigoso do que esconder do espectador as dificuldades reais e fazê-lo acreditar que se pode estudar linguística geral com proveito sem conhecimento de gramática comparada. As consequências dessas miragens são sentidas continuamente no próprio mundo científico e contribuem em grande parte para agravar o cisma entre duas disciplinas que constantemente precisam uma da outra, e para difundir o diletantismo na linguística teórica. Linguística geral sem gramática comparativa é uma linguística vazia.⁴ (Hjelmslev, 1940, p. 124)

A análise do texto de Gray deve ser lida em sua totalidade por qualquer um que queira aprender mais sobre a preparação de Hjelmslev para a tarefa que estava por tomar ao escrever *A Linguagem*. Logo no início da análise, o autor enumera, sem nenhuma pretensão de um levantamento completo, vinte e seis trabalhos gerais sobre linguagem que foram escritos no século dezanove, e reconhece a necessidade que essa

⁴ N. T.: Tradução do original francês: “Un mérite qui sera accueilli sans doute avec une satisfaction particulière est qu’il est tenu compte dans une très large mesure des faits et des théories indo-européens. L’indo-européen reste le stock où le linguiste de tous temps puisera de préférence ses matériaux et ses exemples; en outre l’indo-européen est le seul domaine linguistique que nous *croions* [italics added — FJW] vraiment connaître à fond. Il est donc oiseux de vouloir éviter au lecteur les difficultés techniques de la grammaire comparée; ce serait le tromper et lui dérober l’essentiel; il faut au contraire saisir le taureau par les cornes et faire ce que M. Gray a fait: introduire dans les problèmes de la linguistique générale en les regardant à travers les faits indo-européens. Ce procédé seul permettra de faire comprendre au lecteur que malgré tout la linguistique est une science, sinon dans la théorie, du moins dans la technique, et de le laisser pénétrer, à titre d’observateur, dans le laboratoire du savant. Rien de plus dangereux que de dissimuler au spectateur les difficultés réelles, et de lui faire croire qu’on peut étudier avec profit la linguistique générale sans des connaissances de grammaire comparée. Les conséquences de ces mirages ne laissent pas de se faire sentir dans le monde scientifique même, et contribuent largement à aggraver le schisme entre deux disciplines qui ont constamment besoin l’une de l’autre, et à répandre le diletantisme dans la linguistique théorique. Linguistique Générale sans grammaire comparée est une linguistique en l’air.”

produção acadêmica acelerada procura suprir. Porém as dificuldades para se responder a essa necessidade são enormes:

Os fundamentos da linguística são controversos, hoje mais do que nunca. As explorações teóricas não equivalem à necessidade prática de um tratado global que abarque todos os ramos de nossa ciência de um ponto de vista único e coletivamente adotado. A linguística pode se tornar uma ciência; ainda não o é; as qualidades essenciais de uma ciência, objetividade e precisão, ainda faltam em grande medida. As apresentações de que acabamos de falar trazem inevitavelmente a marca dessa falta; muitas delas sofrem de subjetividade e exclusividade; frequentemente foram escritas para propagar visões particulares, embora amiúde originais; existem algumas delas que são guias francamente infiéis... Entre as muitas apresentações que existem, dificilmente há uma que possa servir como um verdadeiro manual para os não iniciados ou para o iniciante, nem como um guia principal para o leitor informado.⁵ (Hjelmslev, 1940, p. 123)

Mesmo Gray, a despeito do mérito incontestável de seu livro, fora guiado por sua atitude imparcial a uma espécie de ecletismo “qui ne peut pas toujours réussir,” o que leva a conclusão do crítico:

Publicar um livro sobre linguística é um trabalho notadamente ingrato, e a culpa é da linguística e não do autor. A menos que se queira visar o objetivo supremo que consiste em apresentar uma doutrina puramente pessoal e ao mesmo tempo sólida (e em nossa opinião só F. de Saussure conseguiu realizá-lo integralmente) estamos reduzidos a apresentar uma acomodação na qual nem tudo pode estar no mesmo nível.⁶ (Hjelmslev, 1940, p. 126)

Não há dúvida de que quando o próprio Hjelmslev decidiu levar a cabo a “tarefa ingrata”, mesmo que por meio de quadros em miniatura do que ele chamou de pesquisa popular, era ao “objetivo máximo” que ele visava e era o modelo de Saussure

⁵ N. T.: Tradução do original francês: “Les fondements de la linguistique sont controversés, aujourd’hui plus que jamais. Les exploits théoriques n’égalent pas le besoin pratique d’un traité d’ensemble embrassant toutes les branches de notre science d’un point de vue unique et communément adopté. La linguistique peut devenir une science; elle ne l’est pas encore; les qualités essentielles d’une science, l’objectivité et la précision, lui manquent encore dans une très large mesure. Les exposés dont on vient de parler en portent souvent fatalement l’empreinte; beaucoup d’entre eux souffrent de subjectivité et d’exclusivité; souvent ils ont été écrits pour propager des vues particulières, bien que souvent originales; il en est qui sont des guides franchement infidèles... D’entre les nombreux exposés qui existent il n’y a donc guère un seul qui puisse servir comme véritable manuel au non-initié ou au débutant, ni comme guide principal au lecteur averti.”

⁶ N. T.: Tradução do original francês: “Publier un livre sur la linguistique est un travail vraiment ingrat, et c’est la linguistique et non l’auteur qui est en faute. A moins de vouloir viser le but suprême qui consiste à présenter une doctrine purement personnelle et en même temps solide (et à notre avis il n’y a que F. de Saussure qui y ait réussi complètement) on est réduit à présenter un compromis où tout ne peut pas être au même niveau.”

que permanecia em foco. *A Linguagem* é, confessadamente, um livro íntimo que, apesar disso, é firmemente assertivo em sua reivindicação de ser uma obra sólida, livre da subjetividade que Hjelmslev reprova nos demais.

O leitor pode perceber aqui um paradoxo, se não uma contradição, porém Hjelmslev apresenta sua pretensão com confiança, apoiando-se, acima de tudo, em seu cuidado em distinguir entre “o que conhecemos com certeza e o que não conhecemos com certeza” (*hvad vi med sikkerhed véd og hvad vi ikke med sikkerhed véd*). Essa frase, curiosamente, encontra eco na palestra centenária que Hjelmslev apresentou em honra de seu eminente predecessor, Vilhelm Thomsen,⁷ na qual lemos: “Ele desejou distinguir, e distinguiu em todo lugar, o mais precisamente possível, entre, como ele diz, “o que pode ser conhecido” e “o que deve restar apenas como hipótese aproximada”: entre aquilo que se conhece com certeza e aquilo que não se conhece com certeza.” (Hjelmslev, 1942). Essa mesma palestra dá ainda forte indicação acerca de outro ponto no qual Thomsen é referência, ou seja, no que diz respeito ao estilo. Ao discutir o livro de Thomsen, *History of Linguistic Science*, Hjelmslev pontua:

O estilo é despretenhoso, claro e simples, porém é por essa mesma ausência de estilística que se estabelece com uma dignidade extraordinária e uma distinção suave. A elegância reside no conteúdo, somente, nunca na forma; a forma é a mais simples possível, e o seu mérito está precisamente no fato de que ela não se afirma, mas imperceptível, irrepreensível e suavemente provê somente o que é necessário para que o conteúdo aja no ouvinte com toda a sua força meticulosamente calculada. Um tom amistoso e brando perpassa a exposição, na qual o humor subjaz pronto para dar cor quando a ocasião naturalmente surgir. (Hjelmslev, 1942).

Eu não poderia imaginar uma melhor caracterização do estilo de Hjelmslev em *Sproget*.

Mesmo a mais breve discussão sobre o pano de fundo deste livro estaria incompleta sem a menção especial a mais um linguista: Edward Sapir. Na necrologia de Sapir elaborada por Hjelmslev (1939), é o respeito do antropólogo alemão pelos fatos – em conjunto com o uso que ele faz deles – que é enfatizado:

Sapir estava constantemente atento a generalizações apressadas e prejulgamentos sem base; mas seu modo de agir não envolvia o indutivismo estéril; Sapir era um grande visionário e possuidor de uma mente concisa, mas suas teorias gerais eram constantemente verificadas, confirmadas por seu conhecimento factual. O método de

⁷ “Vilhelm Thomsen. Foredrag paa Københavns universitet 26. Januar 1942,” *Gads danske Magasin* (1942), p. 136-47.

Sapir era empírico no melhor sentido da palavra: sua doutrina era a teoria construída pela experiência. O leitor dos trabalhos de Sapir tinha uma sensação constante de que os fatos não eram postos com o intuito de fornecer conhecimento por si só, mas sim para provê-lo com o material necessário para o raciocínio... A teoria da linguagem de Sapir inspira segurança, pois sua confiança nela baseava-se em fatos. Por essa razão, *Language* (1921), de Sapir, deve seguramente manter seu lugar como um estudo clássico e também como uma das melhores introduções à linguística geral já escritas. Quem vos escreve agora lembra-se claramente de seu primeiro contato com a obra; foi uma revelação, uma confirmação da vaga antecipação da possibilidade de estabelecer uma *linguística comparativa geral* destinada a ultrapassar a *filosofia da linguagem*, subjetiva e sentimental, do passado.

Não é surpresa que *A linguagem*, de Sapir (1954 [1921]), encabece a lista de leituras sugeridas adicionadas ao final do livro de Hjelmslev. Por questões de adequação, decidi não reproduzir essa lista na edição em inglês, porém é interessante analisar a lista de obras que inclui trabalhos publicados antes de 1944. Entre as introduções gerais recomendadas, o livro de Sapir é seguido pelo *Le langage*, de Vendryes; *Sprogvidenskaben*, de Sandfeld; *La linguistique*, de Marouzeau; *Philosophy of Grammar*, de Jespersen; *Sprogvidenskaben i det nittende århundrede*, de Pedersen; *Language and Languages*, de Graff; *Language*, de Bloomfield; *Foundations of Language*, de Gray; “Sproget”, de Skautrup (incluso em *Alverdens Viden III*); e *Introduktion i språkventenskapen*, de Collinder. O *Curso de linguística geral*, de Saussure, surge, curiosamente, não entre os trabalhos gerais sobre linguagem, mas nas referências elencadas sob o título “Mudança linguística”, no qual uma menção é especialmente feita ao seu “Mémoire sur le système primitif des voyelles dans les langues indo-européennes”. A única referência sob o título “Formação dos signos” é o *Traité de phonétique*, de Grammont, e a única feita sob “Estrutura linguística e uso linguístico” é *Theory of Speech and Language*, de Gardiner, descrita como “uma apresentação do mesmo problema de um ponto de vista ligeiramente distinto”. O trabalho de Hjelmslev, *Omkring sprogteoriens grundlæggelse*, aparece como leitura adicional sugerida em “Tipos de estrutura linguística”, enquanto *Modersmålets fonetik*, de Jaspersen, e *Kindersprache, Aphasie und Allgemeine Lautgesetze*, de Jakobson, têm seu lugar sob o título de “Tipos de uso linguístico”. O tópico “Relação genética” está representado pela obra *Introduction à l'étude comparative des langues indo-européennes*, de Meillet, *Einführung in das Studium der indogermanischen Sprachwissenschaft*, de Schrijnen, e pelo *Jämförande indoeuropeisk språkventenskap*, de Charpentier. As referências sub a rubrica “Famílias linguísticas” são *Europas och den vita rasens språk*, de Lombard; *Die Sprachstämme der*

Erde, de Kieckers; e *Les Langues du Monde*. As obras *Sprachvergleichung und Urgeschichte*, de Schrader, e “*Olkdarisk kultur*”, de Vilhelm Thomsen (incluso no primeiro volume do *Samlede afhandlinger*), são as leituras recomendadas em “Línguas mães”.

Algumas palavras ainda restam ser ditas sobre a presente versão deste livro. Para além da lista de leitura mencionada anteriormente, eliminei outro apêndice que era mais apropriado à edição original que para a presente versão – uma curta coleção de termos selecionados e suas definições, acompanhados por algumas breves notas biográficas de linguistas mencionados no texto. Como tradutor, não considerei como parte do meu trabalho adaptar o fato de que o livro fora originalmente pensado para um leitor dinamarquês: assim, o leitor ainda encontrará que o Lácio é comparável em tamanho à Zelândia (4345 quilômetros quadrados), e contraexemplos introduzidos por analogia com a situação legal das ilhas Feroe. Ao mesmo tempo, troquei ou modifiquei alguns exemplos ilustrativos que, em razão do conhecimento de dinamarquês que eles pressupunham ou por outras razões, exigiriam a adição de notas de rodapé dispersivas. No campo da terminologia, desviei-me do original em algumas instâncias. Foram usadas “relação” e “correlação” em seus sentidos técnicos empregados nos *PTL*, simplesmente porque não encontrei outra maneira de apresentar, sem risco de causar alguma confusão, o uso menos formal dos termos empregado no *Sproget*. Devo ainda mencionar que, particularmente, o texto em inglês possui elementos da expressão “representados” por sons ou letras, onde o dinamarquês usa palavras que seriam normalmente traduzidas como “nomeadas” ou “significadas”. Acredito que Hjelmslev poderia, em inglês, ter preferido “simbolizado”, mas, ainda com alguns receios, decidi por um substituto relativamente neutro.

O capítulo final, sobre linguagens de diferentes graus, não está incluso na edição dinamarquesa. Entretanto, ele, de fato, constituía originalmente a conclusão do livro, como descobri em 1967 enquanto revia os manuscritos de Hjelmslev sob o gentil convite da esposa de Hjelmslev. Apesar de a passagem dos anos ter obscurecido a história desse capítulo perdido, parece-me que Hjelmslev fora persuadido a acreditar que tal capítulo seria de difícil compreensão para um público leigo e que poder-se-ia reduzir o livro a uma extensão mais razoável caso essa parte fosse retirada. Mesmo que isso seja verdade, o capítulo foi preservado em boa cópia, e claramente representa o resumo do que fora apresentado anteriormente. Por essas razões, não hesitei em incluí-lo aqui. O leitor mais atento pode considerá-lo particularmente útil no sentido de dispersar os

curiosos equívocos acerca da suposta “negligência de Hjelmslev em relação à substância” e da preocupação exclusiva com a forma. Louis Hjelmslev autorizou-me a realizar esta tradução. Ofereço a Vibeke Hjelmslev com minha mais calorosa gratidão.

F. J. W.

Berkeley

3 de outubro de 1969

Observações preliminares

A ciência da linguagem, ou linguística (do Latim *lingua* “língua”), teve – como qualquer outra ciência – seus períodos clássicos e também críticos. Nos períodos clássicos, havia um corpo (ou estrutura) estabelecido da doutrina, acordado e seguido por todos; e o trabalho científico dedicava-se aos detalhes que poderiam ser alocados nos quadros de tal estrutura. Nos períodos críticos, investigadores ocuparam-se com a estrutura em si, que tentaram estabelecer sobre novas ou melhores bases, refletindo uma apreciação mais profunda e perscrutante da natureza da linguagem.

A linguística talvez nunca tenha passado por um período crítico de tal profundidade e alcance como o de nosso tempo. A própria natureza da linguagem e as peculiaridades estruturais estão sendo agora debatidas por linguistas. A própria base da linguística está sendo reforjada. Isso não é dito, naturalmente, com o objetivo de insinuar que os linguistas tenham parado de se preocupar com os detalhes. Mas, mesmo no seu trabalho com os detalhes, eles se veem constantemente diante de problemas basilares de um nível muito mais alto. Essa preocupação com os fundamentos da linguística já nos levou tão além das primeiras noções dessa ciência que a nova linguística de nosso período pode ser contrastada com todas as variedades prévias, que podem ser agrupadas sob o mesmo nome de *linguística clássica*. Mais cedo ou mais tarde a presente reelaboração deverá necessariamente levar a um novo classicismo. Talvez possamos mesmo agora dizer que chegamos longe o suficiente para elaborar ao menos os contornos de um tal novo classicismo, mas ainda não o alcançamos e não existe também um assentimento sobre os caminhos e os meios para esse fim.

Uma crise desse tipo não deve ser lamentada – muito pelo contrário. Os períodos críticos são aqueles renovadores e inspiradores, que provam que a ciência não está estagnada, mas em constante renovação de si mesma. Tais períodos alargam nosso horizonte e aprofundam nosso entendimento das coisas mais importantes. De fato, tão forte tem sido o efeito da atual crise que podemos até arriscar afirmar que apenas agora a linguística conseguiu se organizar como uma ciência independente.

É possível interessar-se por linguagem a partir de muitos pontos de vista diferentes. Os Gregos e Romanos da antiguidade e os pensadores da escolástica medieval estavam interessados particularmente a partir de um ponto de vista lógico; e foi em uma associação estreita com a lógica aristotélica que fundaram uma tradição gramatical que ainda sobrevive na nossa gramática escolar e, com poucas mudanças, ainda na maior parte

da nossa gramática científica em suas formas clássicas. Nos séculos dezoito e dezenove, muitos estudiosos interessaram-se por linguagem principalmente de um ponto de vista histórico e pré-histórico, com uma referência particular à história cultural, de tal forma que a história e as relações entre línguas eram vistas como um reflexo, e um meio de conhecimento, da história e das relações entre os povos. Ascendeu durante o século dezenove a ciência que veio a se chamar de fonologia ou fonética, interessada na linguagem de um ponto de vista fisiológico e físico e que procurava descrever o plano da expressão da língua por meio de um estudo dos movimentos e posições dos órgãos fonadores na produção dos sons da fala e por meio de um estudo das ondas sonoras que são assim produzidas. Também ganhou adeptos durante o século dezenove uma linguística psicológica, que viu no lado psicológico da linguagem, aqui tida como parte da “vida mental” do ser humano, a característica essencial da natureza da linguagem. De volta ao nosso próprio tempo, as considerações sociológicas, de maneira cada vez mais crescente, têm sido colocadas em cena para sustentar a linguagem como uma instituição social.

Evidentemente, não faltam pontos de vista a partir dos quais a linguagem pode ser, e tem sido, estudada. Porém nenhum desses (e outros poderiam ser facilmente elencados) fornece as bases para uma ciência da linguagem independente; pelo contrário, a linguagem se torna um objeto de estudo ora da lógica, ora da história, ora da fisiologia, física, psicologia ou sociologia. Ainda se pode argumentar que, apesar de todos esses estudos multifacetados, um ponto de vista sobre a linguagem tem sido negligenciado, justamente aquele que parece o mais importante e o mais natural – o ponto de vista *linguístico*. Deve ser possível imaginar uma ciência que não tome a linguagem apenas como um conglomerado de fatos lógicos, históricos, fisiológicos, físicos, psicológicos e sociológicos, mas primordialmente e, acima de tudo, como uma entidade independente, uma formação integral de um tipo especial. Apenas uma tal visão integral pode dar conta do fato de que todos esses elementos, aparentemente tão heterogêneos, são capazes de se conjuntar na linguagem. O ponto de vista linguístico, tomado como central, pode determinar a justificação relativa e a limitação relativa de cada ponto de vista particular. Dessa forma, apenas quando as linguísticas lógica, histórica, fisiológica, física, psicológica e sociológica forem complementadas por uma linguística *linguística*, uma ciência da linguagem poderá ser estabelecida. É em um esforço com esse intuito que nossa presente crise nos encontra.

Qualquer apresentação que seja feita atualmente sobre o problema da linguagem deve, necessariamente, refletir essa situação – tanto para o bem quanto para o mal. Uma vantagem que deriva desse debate crítico é que acreditamos ter atingido um claro entendimento de muitas coisas que eram antes simplesmente ignoradas ou vistas sob uma ótica bastante distinta. Uma desvantagem encontra-se na multiplicidade de vozes discordantes que ecoam ao redor do autor enquanto ele escreve. Dificilmente seria possível, em um estudo de cunho introdutório como este, permitir que todas essas vozes sejam ouvidas e introduzidas diretamente no debate crítico. Ademais, seria de pouca serventia. Existem muitos pontos de vista – de fato, muitos importantes – nos quais encurta-se o debate e se diz: as coisas são dessa forma e não dessa outra forma. Seria, então, uma confusa perda de tempo explicar que existem pessoas que pensavam, ou que talvez ainda pensem, que as coisas são diferentes. Muitas posições foram defendidas sem nenhum tipo de justificativa no passado subjetivo da linguística. Através da presente crise, demandam-se justificativa, argumentação e provas; opiniões dadas sem justificativa não devem ser imputadas ou refutadas: elas morrerão por si mesmas.

O leitor deve perceber, por outro lado, que dificilmente seria possível estabelecer um sistema completo no presente momento. Em diversos aspectos, a apresentação que se segue deve ter como objetivo colocar problemas mais do que resolvê-los. Este livro, portanto, foi pensado como uma introdução ao *problema* da linguagem atualmente, segundo o *juízo do autor*.

O leitor deve também compreender duas coisas:

1. Nosso estudo *não é subjetivo*. É claro que um forte elemento pessoal é inevitável em relação aos fins e aos meios da pesquisa, bem como na maneira pela qual os problemas são concebidos, mas as perspectivas que são oferecidas aqui não são subjetivas: são objetivas, baseadas em argumentos. Elas representam, não mera crença, mas um pensamento substanciado pela razão.

2. Nosso estudo *não é exclusivo*. Tenta-se fazer justiça a todos os campos da linguística. O fato de que a ciência está adotando novos pontos de vista não quer dizer, como um leigo poderia ocasionalmente acreditar, que, por essa razão, abandonar-se-ão todos os resultados previamente obtidos, tomando-os como dispensáveis ou até mesmo lamentáveis. Mesmo que Tycho Brahe erroneamente tenha acreditado que o sol girava em torno da terra, isso não significa que suas observações, descobertas e resultados devem ser abandonados, mas apenas que devem ser *reinterpretados*, analisados de acordo com

as exigências do ponto de vista atual e, assim, entendidos de uma maneira diferente. O mesmo vale para os campos dentro da linguística clássica nos quais verdadeiras observações e descobertas foram feitas. Em ciência, é certamente possível falar de resultados permanentes – embora dificilmente de pontos de vista permanentes. A linguística clássica do século dezanove evidentemente chegou a resultados permanentes a respeito das relações genéticas entre as línguas, resultados esses que constituem um campo essencial da linguística. Tais resultados serão apresentados aqui, porém ajustados a novos pontos de vista, o que demandará um manejo um tanto diferente daquele no qual eles provavelmente foram vistos à época de suas descobertas; assim, esta apresentação diferirá em seus princípios daquilo que é usualmente encontrado nos manuais sobre esse assunto.

Uma exposição ajustada dos principais resultados da linguística do século dezanove será nosso ponto de partida, pois o campo ao qual pertencem – o estudo das relações genéticas entre as línguas – ainda é o mais bem desenvolvido no interior de toda a ciência linguística e, também, porque é o único para o qual a linguística clássica produziu o que podemos chamar – ao menos depois de uma reinterpretação – de método exato. Essa pedra angular do classicismo do século dezanove é também o que permitiu a ascensão da abordagem crítica da linguística moderna e será utilizada dessa forma com o objetivo de introduzir o leitor à perspectiva moderna dentro da qual os problemas da estrutura linguística e da mudança linguística serão tratados.

Função linguística

atingimos discernimento, familiarizamo-nos, entendemos ou compreendemos a linguagem da mesma forma que assimilamos outros objetos – por meio de uma descrição. E descrever um dado objeto pode significar apenas uma coisa: dar conta das relações das quais faz parte ou daquelas que o constituem. A essas relações, ou dependências, registradas por uma descrição científica, daremos o nome de *funções*. Assim, podemos descrever um dado objeto de duas maneiras: (1) dividi-lo em partes com função mútua, por meio de uma *análise*; (2) colocá-lo em conjunto com um grupo cujas partes possuem função mútua, por meio de uma *síntese*. No primeiro caso, o objeto é visto como um todo funcional; no segundo, é visto como parte de um todo funcional maior.

Consequentemente, quando o objeto que queremos descrever é uma língua, podemos:

1. por um lado, analisar a língua em partes com função mútua; isso é feito pelo campo da linguística conhecido como *gramática*, que vê a língua particular (mais precisamente o *état de langue* particular) como uma totalidade fechada e a descreve por meio de uma análise que leva em conta as funções entre suas partes;

2. por outro lado, colocar a língua dentro de um grupo maior, ou seja, vê-la como parte de um *grupo de línguas* e, então, descrever tal classe por meio de uma análise que dá conta das funções entre suas partes.

As conexões entre línguas pertencentes a uma e mesma classe de línguas – ou, em outras palavras, as funções que estabelecem uma classe linguística – constituem uma *relação linguística*, da qual distinguimos dois tipos:

1. relação *genética*, encontrada entre línguas que pertencem à mesma *família linguística* e que aponta para uma origem em comum para elas.

2. relação *tipológica*, encontrada entre línguas que pertencem ao mesmo *tipo linguístico*; uma relação tipológica não aponta para uma origem em comum, mas baseia-se em um acordo de características estruturais que é condicionado pelas possibilidades gerais da linguagem.

No que se segue, veremos que tipos de funções estabelecem esses dois tipos de classes de línguas. Teremos também a oportunidade de determinar algumas das funções mais importantes do *état de langue*. Na verdade, é claro que a análise do *état de langue* particular é um pré-requisito para a classificação das línguas, mas o problema principal da linguística sempre será não o *état de langue* individual, mas a relação entre

diferentes estágios de uma mesma língua e entre diferentes línguas, suas similaridades e suas diferenças. Esse problema central será, portanto, abordado diretamente, enquanto as questões concernentes ao état de langue serão discutidas conforme apareçam e requirem elucidação ao longo da exposição.

Relações genéticas entre as línguas

Nossa própria língua pertence à família conhecida comumente como *indo-europeu* (assim chamada porque engloba línguas tanto da Índia como da Europa). Dessarte, alguns exemplos de palavras dessa família serão apresentados para ilustrar o conceito de relação genética. Se tais exemplos foram escolhidos da família indo-europeia e não de alguma outra, como também poderiam ter sido, isso não é somente porque a família indo-europeia inclui línguas particularmente bem conhecidas (apesar de o conhecimento das línguas citadas aqui por parte do leitor não ser importante para o entendimento do argumento) e porque nossa língua pertença a ela, mas porque, de todas as famílias linguísticas, essa tem sido a mais profundamente estudada e seu percurso traçado através do mais longo período de evolução e da mais ampla difusão geográfica. Nosso conhecimento das línguas indo-europeias se estende por um período de mais de 3500 anos, e as línguas indo-europeias são agora faladas por 1,4 bilhões de pessoas, aproximadamente metade da população mundial.⁸ Ao escolher essa família como exemplo, podemos demonstrar que nosso método não está sujeito a quaisquer limitações de tempo ou de espaço.

Por outro lado, devemos escolher os exemplos dentre os muitos disponíveis de modo a apresentar nosso material na forma mais simples possível e também para evitar complicações desnecessárias que podem, à primeira vista, ofuscar o quadro geral. De maneira a satisfazer esses dois requisitos – demonstração sobre uma grande extensão de tempo e espaço, e simplicidade da exposição – escolheremos nossos exemplos de certas línguas indo-europeias que (por razões que poderão ser discutidas mais tarde) são particularmente apropriadas para esse fim. A escolha das línguas variará de um exemplo para o outro, mas na maioria dos casos usaremos as línguas listadas no parágrafo seguinte.

Dentre as línguas que possuem uma relação genética mais aproximada com a nossa própria,⁹ escolhemos o *gótico* (gót.), falado desde o ano 200 até aproximadamente o ano 500 em vastas áreas da Europa, da Espanha no Oeste até à Criméia, no Leste, e na qual a Bíblia fora traduzida no século quatro, onde hoje se encontra o território búlgaro. Somente quando uma palavra não é encontrada em gótico (isto é, não é identificada em

⁸ N. T.: Atualmente, segundo dados do site Ethnologue (2019), o total de falantes de línguas indo-europeias aproxima-se de 3,3 bilhões. O que ainda corresponde à metade da população mundial.

⁹ N. T.: Como já fora salientado, a centralidade das línguas germânicas na exposição se dá por conta do parentesco com o dinamarquês.

textos góticos) é que escolheremos outras línguas estreitamente relacionadas em seu lugar: *nórdico antigo* (n.a.), a língua das sagas islandesas e das Eddas norueguesas; *alto-alemão antigo* (a.a.a.); ou *inglês antigo* (i.a.). Para representar as línguas celtas, usaremos ou *irlandês* (ir.) ou *galês* (ga.). Ademais, citaremos *latim* (lat.), *grego* (gr.), *lituano* (lit.), *antigo eslavo eclesiástico* (a.e.e.), *armênio* (arm.) e *hindi antigo* (h.a.) – e, em alguns casos, *hitita* (hit.), uma língua antiga da Ásia menor, conhecida a partir de um grande número de inscrições cuneiformes, e *tocariano* (toc.), uma língua morta, preservada em manuscritos encontrados no Turquestão oriental. No momento, não vemos necessidade de informações adicionais sobre essas ou outras línguas relacionadas, porém mais informações serão dadas no capítulo sobre as famílias linguísticas.

Veremos agora, primeiramente, as palavras *irmão* (ing. brother)¹⁰, *mãe* e *pai*.

A palavra *irmão* é citada aqui a partir de todas as línguas citadas acima, com exceção apenas ao hitita:

gót. <i>broþar</i>	gr. <i>phrátōr, phrátēr</i>	arm. <i>etbayr</i>
ir. <i>bráthir</i>	lit. <i>broter-ėlis</i>	i.a.. <i>bhrátā</i>
lat. <i>frāter</i> (cf. Eng. fraternize)	a.e.e. <i>bratrъ</i>	toc. <i>pracarə</i>

Observações:

1. Até o momento, no que se refere às letras das transcrições, é suficiente, para o nosso propósito, que o leitor seja capaz de reconhecê-las quando as encontrar novamente. Além disso, salientamos que o leitor não deve se preocupar com essas particularidades ou tentar imaginar a pronúncia das palavras elencadas. As notas que vêm a seguir podem servir para sanar essa curiosidade que é, entretanto, natural: *þ* é pronunciado como *th* em ing. *thing*¹¹; um traço horizontal sobre uma vogal indica que esta é longa; *ł* é pronunciado como um *l* mas com a parte posterior da língua levantada (*l* velarizado)¹²; *ə* é pronunciado como *a* em ing. *sofa*;¹³ AEE *ъ* é uma vogal extra curta; e o til (~) indica uma entonação especial; O *c* em toc. é pronunciado como *tch* em *tchau*, em português.

¹⁰ N. T.: A palavra “irmão” em português não tem a mesma origem da palavra nas demais línguas atestadas. Quando esse for o caso, recorreremos ao inglês como referência.

¹¹ N. T.: Trata-se de uma consoante fricativa linguodental surda. Em português não encontramos pronúncia similar.

¹² N. T.: Essa realização de *l* é chamada “*l* escuro” em inglês ou “*l* duro” em russo. O *l* velarizado acontece em final de palavra em certos dialetos do Sul do Brasil, em palavras como “sul” e “sal”.

¹³ Trata-se de vogal central frouxa, encontrada em grande parte das vogais inacentuadas em inglês e francês (ver a pronúncia do primeiro “e” em “tristement”). Apesar de não aparecer no português brasileiro, essa

2. A forma lituânea é um hipocorístico e, dessa maneira, suprida pelo sufixo *-ėlis*, o qual não encontramos nas palavras citadas de outras línguas. Outra palavra para “brother” no Lituano é *brólis* (na qual o acento agudo indica uma entonação especial que é diferente daquela com til), que, apesar de relacionada à palavra *irmão*, não é sinônima desta.

3. A palavra em questão significa “irmão” em todas as línguas citadas, exceto na literatura grega, no qual possui um significado especial “membro do (mesmo) clã”, mas sabemos que essa mesma palavra no dialeto iônico (que possui a forma ligeiramente diferente *phrētēr*) significava, sim, “irmão”. No grego comum, “irmão” é *adelphos*, que é uma palavra diferente. Um “clã”, isto é, uma “irmandade”, em grego é *phrātría*, um derivativo da palavra *irmão* que pode também ser encontrada no a.e.e. *bratrija* e no h.a. *bhrātryam*.

A palavra *mãe* não foi preservada no gótico (na bíblia gótica, “mãe” aparece como *aiþei*, que deve ser uma palavra diferente, porém podemos elencar:

n.a. <i>móðir</i>	gr. <i>mātēr</i> (Ático <i>mētēr</i>)	arm. <i>mayr</i>
ir. <i>máthir</i>	lit. <i>mótė</i> , <i>motė</i> “mulher”, <i>mótina</i> “mãe”	h.a. <i>mātá</i>
lat. <i>māter</i>	a.e.e. <i>mati</i>	toc. <i>mācarə</i>

(A não ser que seja especificado de outra forma, esta palavra, em cada língua citada, significa “mãe”. Pronúncia: ð ocorre como *th* no ing. *there*;¹⁴ Ir. *th* como em ing. *thing*.)

A palavra *pai* não é encontrada em lituano ou em antigo eslavo eclesiástico: a palavra lituana para “pai” é *tėvas*, que pode ter algum grau de parentesco distante com nosso exemplo *pai*, porém de qualquer forma, não é idêntica a esta; a palavra em antigo eslavo eclesiástico é *otьcь*, uma palavra completamente diferente. De qualquer maneira, podemos citar as seguintes palavras, todas tendo como significado “pai”:

gót. <i>fadar</i>	toc. <i>pācarə</i>
ir. <i>athir</i>	gr. <i>patēr</i>
lat. <i>pater</i> (cf. Ing. <i>paternity</i>)	arm. <i>hayr</i>
	h.a. <i>pitā</i>

Voltaremos nossa atenção agora para os *elementos da expressão* que compõem essas palavras nas diferentes línguas apresentadas. Tais elementos da expressão são

vogal apresenta-se também em sílabas não acentuadas do português europeu (ver a pronúncia do “a” em “cadeira”).

¹⁴ N. T.: A diferença de *th* para ð é que este último é uma consoante sonora.

representados na escrita pelas letras, e na língua falada pelos sons da fala (em alguns casos, usamos uma combinação de letras para representar apenas um elemento da expressão: ir. *th*, gr. *ph* e h.a. *bh* serão contabilizados como a representação de um elemento da expressão por cada). Se não levarmos em consideração os acentos (que, é claro, são também elementos expressivos), as palavras gregas *phrátēr*, *mâtēr* e *patēr*, por exemplo, serão analisadas como possuindo seis, cinco e cinco elementos expressivos, respectivamente.

Neste ponto podemos então demonstrar que, entre cada elemento da expressão (EE) em uma língua e cada EE em outra língua geneticamente relacionada, existe uma dependência ou relação, ou, como acordamos chamá-la, uma função. É, de fato, nessa função entre os EE nas diferentes línguas – essa função-elemento – que repousa a relação genética entre as línguas. Incidentalmente, é também por causa dessa função-elemento que encontramos recorrentemente, em exemplos como os vistos anteriormente, uma semelhança entre palavras em diferentes línguas (apesar de que essa semelhança não precisa estar presente – seria difícil afirmar que as palavras em armênio trazem consigo uma “semelhança” muito grande com as de outras línguas).

Um caso especialmente simples é o primeiro EE na palavra *mãe*, que aparece como um *m* em todas as línguas que citamos acima. O que é interessante e crucial é a possibilidade de se estabelecer uma regra que diga que a cada vez que temos um *m* em início de sílaba em uma palavra de uma dessas línguas, temos um *m* na mesma palavra em todas as outras línguas, contanto que a palavra seja encontrada nelas. Para não haver equívocos, tal regra não deixa de ter algumas restrições, mas essas restrições podem ser estabelecidas de forma definitiva e mantidas em separado. Retornaremos a essa questão mais tarde, em nossa discussão sobre a formação dos signos. Antes de tudo, demonstraremos a precisão de nossa regra por meio de alguns exemplos:

rato: n.a. *mús*, lat. *mūs*, gr. *mūs*, arm. *mukn*, h.a. *mūk*.

lua, com o seu derivativo, *mês*: gót. *Mena* “lua”, *menoþs* “mês”; ir. *mí* “mês”; lat. *mēnsis* “mês”; gr. *mén* “mês”, *ménē* “lua”; lit. *ménuo* “lua”, *ménesis* “mês”; a.e.e. *měšęčB* “lua, mês”; arm. *amis* “mês”; h.a. *māh* “lua, mês”. (o *ě* do a.e.e. é pronunciado aproximadamente

como *a* no ing. *man*;¹⁵ *ɛ* como no fr. *fin*;¹⁶ *c* como *ts*; *B* é uma vogal extra curta que é diferente de *ɛ*; o *h* h.a. é apenas um *h* comum derivado de *s*.)

me (mim): gót. *mi-k*, ga. *mi*, lat. *mē*, gr. *me*, *emé*, *emé-ge*, lit. *manė*, a.e.e. *mę*, h.a. *mā*.

morte: gót. *maurþr* “assassinato”; lat. *mors* “morte”; gr. *mortós* “mortal”; lit. *mirtis* “morte”; a.e.e. *sъ-mrъtъ* “morte”; arm. *mard* “um homem” (um mortal); h.a. *mṛtiḥ* “morte”. (*r* indica um *r* silábico).

inverno: n.a. *gimbr* “cordeiro com um ano (um inverno) de idade” (cf. ocorrência dialetal em ing. *gimmer*). ir. *gemred* “inverno”; lat. *hiems* “inverno”; gr. *kheima* “geada de inverno”; lit. *žiemà* “inverno”; a.e.e. *zima* “inverno”; arm. *Jmeṛn* “inverno”; h.a. *himáh* “geada, neve”, *himā* “inverno”, *héman* “no inverno”; hit. *kimanza* “inverno”. (*ž* é pronunciado como *j* no port. *jogo*; hit. *z = ts*; arm. *j = dz*; arm. *ṛ* é um forte *r* apical.¹⁷)

domar: gót. *ga-tamjan* “domesticar”; n.a. *tamr* “domar”, *temja* “domesticar”; ir. *damnaim* “eu amarro, eu domo”; lat. *domāre* “domesticar”; gr. *damân* “domesticar”; h.a. *dāmyati* “ele doma, é domado”.

Quando *m* está no final de uma sílaba, ele novamente surge como um *m* na maior parte das línguas, mas no antigo eslavo eclesiástico ele se combina com a vogal precedente para produzir uma vogal nasalizada. Começando pelo significado “serrilhado”, podemos identificar a palavra *comb* (pente), n.a. *kambr* “pente, crista, praia recortada”, com o gr. *gómphos* “pino, cavilha”, h.a. *jámbhaḥ* (*j* é pronunciado como o *d* em Port. *dia*),¹⁸ lit. *žam̃bas* “borda irregular”. No antigo eslavo eclesiástico, a mesma palavra é *zqbъ* “dente”. Este vai ser o caso em todos os exemplos nos quais *m* surge ao final de uma sílaba interna. A relação é novamente outra quando *m* aparece no final de uma palavra: nesse caso, em Lituano, encontramos uma vogal nasal (que não é mais pronunciada como nasal na língua padrão); no antigo eslavo eclesiástico, por vezes encontramos nasalização da vogal, outras vezes o zero, a depender de certas condições da prosódia; no grego, encontramos *-n*; em gótico e em nórdico antigo encontramos ocorrência da realização zero. Dessa forma, a palavra que nós citamos acima no caso nominativo irá assumir, na forma acusativa (no indo-europeu em palavras terminadas em

¹⁵ N. T.: Trata-se de vogal oral anterior mais baixa que o “é” em “pé” e mais alta que o “a” em “chá”, em português.

¹⁶ N. T.: Trata-se de uma vogal anterior meio-aberta não-arredondada nasal. Não há correspondente em português, que só nasaliza a vogal anterior meio-fechada.

¹⁷ N. T.: Em português, realiza-se com uma vibrante múltipla, encontrada em algumas variantes regionais do sudeste brasileiro. (BARROS, 2017).

¹⁸ N. T.: Para os dialetos que transformam a oclusiva dental sonora *d* em consoante africada antes de *i* e suas variantes.

-m), as seguintes formas: h.a. *jámbham*, lit. *žam̃bq*, a.e.e. *zqb̃*, gr. *gómphon*, n.a. *kamb*. Se a palavra tivesse existido também em Latim, a forma acusativa terminaria em *-m*, assim como as outras formas acusativas.

Dessarte, encontramos uma função constante entre os EEs dessas línguas, de tal forma que as mesmas correspondências surgem sempre sob as mesmas circunstâncias. Ao usar o símbolo φ para designar essa dependência ou função constante, podemos afirmar que na posição inicial da sílaba teremos:

gót. $m \varphi$ celt. $m \varphi$ lat. $m \varphi$ lit. $m \varphi$ a.e.e. $m \varphi$ arm. $m \varphi$ h.a. m .

Porém seria demasiadamente trabalhoso escrever ou dizer tudo isso sempre que quisermos nos referir a tal função-elemento entre todas as línguas indo-europeias e, por essa razão, usaremos um sistema mais conveniente no qual uma única função que permeia todas as línguas é designada por um único símbolo. Naturalmente, esse símbolo – como todos os símbolos científicos – é escolhido de maneira totalmente arbitrária. Poderíamos, caso quiséssemos, acordar em utilizar x ou y ou z para referimo-nos à extensa fórmula funcional acima. Entretanto, os símbolos utilizados nas fórmulas científicas não são escolhidos somente de forma arbitrária, mas também apropriada e, assim sendo, a partir de considerações exclusivamente práticas, escolheremos uma fórmula que nos lembrará diretamente algo daquilo que se pretende designar. No exemplo em pauta, não há dúvida de que o mais simples será escolher a letra m , antes da qual introduziremos um asterisco para indicar que ela é uma fórmula de um elemento-função. Assim, ao invés de dizer sempre “concordância entre o m gótico, o m céltico e o m latino, etc., no início de uma sílaba,” diremos “ m indo-europeu” e escreveremos “IE * m .”

Para além da arbitrariedade e da adequação, as fórmulas científicas ainda possuem a propriedade adicional da simplicidade. Quando decidimos estudar todos as funções-elemento entre as línguas indo-europeias e encarar a tarefa de registrá-las, é preferível utilizar o menor número possível de símbolos. Ainda em nosso mesmo exemplo, vemos que as condições sob as quais a função é diferente (posição final de sílaba e final de palavra) e as condições sob as quais nossa primeira função se mantém (posição inicial de sílaba) são mutuamente excludentes. Dessa forma, podemos nos permitir o uso do mesmo símbolo em todos esses casos sem correr o risco de confusão. E já que existem línguas nas quais o EE analisado é representado por m nas posições final de sílaba e final de palavra, podemos também utilizar o símbolo * m e estabelecer as seguintes equações:

I.E. $*m =$ gót. n.a $m \varphi$ gr. $m \varphi$ lit. $m \varphi$ a.e.e., φ h.a. m

em posição final de sílaba no interior da palavra, e

I.E. $*-m =$ gót. n.a. $\emptyset \varphi$ lat. $-m \varphi$ gr. $-n \varphi$ lit. $, \varphi$ a.e.e. $\emptyset (,) \varphi$ h.a. $-m$.

Na última fórmula, fomos capazes de estabelecer a condição (posição final de palavra) simplesmente colocando um hífen antes do m .

Estabeleceremos a mesma organização para todos os EEs, um por um, lembrando-nos sempre de utilizar o menor número possível de fórmulas indo-europeias e, conseqüentemente, de reunir sob a mesma fórmula o máximo possível de condições diferentes. O método é puramente mecânico e facilmente aplicável contanto que o material léxico esteja disponível. Assim, o elemento final em lat. *frater*, *mater* e *pater* pode ser identificado com o segundo elemento em *frater* e denotado pela fórmula $*r$. O leitor pode observar facilmente a partir dos exemplos – aos quais muitos outros podem ser adicionados – que a maioria das línguas sob a maior parte das condições possui um r correspondente, apesar de o lituano e o hindi antigo, em posição final de palavra, possuírem o zero. Esse zero será, certamente, substituído pelo r quando mudarem as condições – isto é, quando forem introduzidas na mesma palavra, o quanto possível, as condições sob as quais a língua objeto possuir um r correspondente: no acusativo, o hindi antigo possui *brhátar-am*, *mātár-am*, *pitár-am* e o lituano possui *móter-į*; cf. também Lit. *broter- ėlis*.

Os mesmos exemplos (e, novamente, muitos outros podem ser elencados) tornam imediatamente óbvio que o terceiro EE em *brother* e o segundo EE em *mother* são idênticos e podem ser formulados como:

I.E. $*\bar{a} =$ gót. o , n.a. $ó \varphi$ ir. $á \varphi$ lat. $\bar{a} \varphi$ gr. $\bar{a} (\bar{e}) \varphi$ lit. $o \varphi$ a.e.e. $a \varphi$ arm. $a \varphi$ h.a. $\bar{a} \varphi$ toc. $\bar{a} (a)$.

No I.E. $*a$ pode também aparecer no final de uma palavra, porém sob essa condição, ele corresponderá a outras fórmulas funcionais. Já apresentamos exemplos desse tipo: sendo $*-\bar{a}$ uma terminação nominativa feminina (cf. a primeira declinação do latim, em $-a$), é encontrada em um grande número de palavras incluindo aquelas que

citamos acima em h.a. *himā* “inverno”, lit. *žiemà*, a.e.e. *zima* e gr. *ménē* “lua”, que seria *ménā* em dialetos que não o ático e o iônico. A partir disso podemos concluir que:

I.E. *-ā = lat. -a φ gr. -ā (-ē) lit. -a φ a.e.e. -a φ h.a. -ā.

O segundo EE em lat. *pater* revela a função:

gót. a φ ir. a φ lat. a φ gr. a φ arm. a φ h.a. i φ toc. ā,

que podemos resumir no símbolo I.E. *_A (reservando **a* para os casos nos quais h.a. *a* corresponda ao *a* das outras línguas, tal como em h.a. *ájati* “dirige”, arm. *acem* “eu lidero, levo”, gr. *ágein* “liderar”, lat. *agere* “dirigir, liderar”, ir. *agat* “deixá-los liderar”, n.a. *aka* “dirigir”). Essa correspondência permanece verdadeira somente para a primeira sílaba de uma palavra; outras sílabas apresentam outras funções-elemento correspondentes a IE *_A.

Iremos concluir nossa amostragem de funções-elemento a partir de uma visada mais aproximada do primeiro EE na palavra lat. *pater* e no terceiro EE – a partir do último – das palavras em lat. *frater*, *mater* e *pater*. Para o primeiro, escreveremos I.E. **p*, e para o último, I.E. **t*.

A palavra lat. *pater*, então, demonstra a seguinte função

I.E. **p* = gót. *f* φ ir. Ø φ lat. *p* φ gr. *p* φ arm. *h* φ h.a. *p* φ toc. *p*,

Outra função que encontramos também em um número enorme de outros exemplos, alguns dos quais também demonstram que o correspondente Hitita é *p*. Citamos abaixo alguns deles:

pena: n.a. *fjōðr*, a.a.a. *fedara* “pena”; a.a.a. *fed(a)rah* “asa”; ir. *én* “pássaro”; galês antigo *eterin* “pássaro”; lat. *petere* “procurar, aspirar a, viajar, partir”; gr. *péteshai* “voar, precipitar-se”; h.a. *pátati* “voa”, *pátram* “asa”; hit. *pítaizi* “foge”, *pítar* “asa”.

peixe: gót. *fisks*, ir. Médio *iasc*, lat. *piscis* (em eslavo, pol. *piskorz* “botia” - um tipo de peixe, russ. *piskar* “gobião” - outro tipo de peixe). (lat. *c* é pronunciado como *k*.)

i.m. *fee* “gado, propriedade”: gót. *faihu* “propriedade, dinheiro”; n.a. *fé* “gado, dinheiro, propriedade”; lat. *pecu*, *pecus* “gado”, *pecūnia* “riqueza, dinheiro”; h.a. *pásu*, *paśú*, *paśúh* “gado”.

din. (*i*)*fjor* “ano passado”: n.a. *fjōðr*, gr. (dórico) *péruti*, lit. *pérnai*, arm. *heru*, h.a. *parút*.

As condições para essa função acabam sendo um tanto quanto restritas, quando consideramos todas as línguas que dela participam. Em armênio, por exemplo, não encontramos *h*, mas sim o zero, quando **p* está na posição de início de palavra antes de I.E. **o* (assim, arm. *otn* corresponde ao ing. *foot*, got. *fotos*, gr. *poús*). Porém sob a maioria das circunstâncias, essa função é encontrada no início das palavras. No grupo **sp*, encontramos *p* em todas as línguas, incluindo-se aqui o gótico; então, por exemplo, n.a. *spá* “prever”, a.a.a *spehōn* “espionar”, quando comparado com o lat. *specere*, *spectāre*, *īn-spicere* “inspecionar”, *haru-spex* “augúrio (aquele que *inspeciona* as entranhas da vítima em busca de presságios)”. Porém, no interior da palavra, **p* corresponde a funções bem diferentes, como veremos mais à frente.

O EE da função got. n.a. *f* \varnothing lat. gr. *p* foi reconhecido primeiramente por Rasmus Rask, em 1814 (em seu ensaio premiado intitulado *Investigation on the Origin of the Old Norse or Icelandic Language*), em conexão com todo um conjunto de outras funções de EE de posição inicial de palavra nessas mesmas línguas, como o gót. n.a. *t* \varnothing lat. gr. *d*, dos quais vimos um exemplo na palavra *domar*. Para esta última função (e também para as funções correspondentes sob outras condições), estabelecemos a fórmula I.E. **d* e, a partir do exemplo de *domar*, complementado pelo material apropriado do lituano, antigo eslavo eclesiástico e armênio, podemos formular a seguinte função:

I.E. **d* = gót. n.a. *t* \varnothing ir. *d* \varnothing lat. *d* \varnothing gr. *d* \varnothing lit. *d* \varnothing a.e.e. *d* \varnothing arm. *t* \varnothing h.a. *d*.

Encontramos bons exemplos para essa função nos numerais *dois* e *dez*:

gót. *twai*, ir. *dáu*, lat *duo*, gr. *dúo*, lit. *dù*, a.e.e. *dьva*, h.a. *dvau*.

gót. *taihun*, ir. *deich*, lat. *decem*, gr. *déka*, lit. *dėšimt*, a.e.e. *desętь*, arm. *tasn*, h.a. *dása*. (š é pronunciado como no port. *ch*.)

O conjunto de funções de EE em posição inicial de palavra identificados por Rask também inclui a função para **t*: gót. n.a. *þ* \varnothing lat. gr. *t*, ou, mais integralmente:

gót. n.a. *þ* \varnothing celt. *t* \varnothing lat. *t* \varnothing gr. *t* \varnothing lit. *t* \varnothing a.e.e. *t* \varnothing arm. *th* \varnothing h.a. *t*,

como exemplificado em:

sede: gót. *þaurus* “seco”, *þaurstei* “sede”, *þaurusjan* “ter sede”, *ga-þairsan*, *ga-þaurusnan* “secar”; ir. *tart* “sede”; lat. *torridus* “seco”, *torrēre* “secar, assar”; gr. *térsesthai* “tornar-se seco”, *tersáinein* “secar”, *trasiá*, *tarsiá*, *tarsós* “caixote de secagem”; arm. *thar* “haste para secar uvas”; h.a. *trşyati* “está sedento”, *tarsah* “sede”, *trşıáh* “seco, áspero”. (*ş* e *t* são pronunciados com a ponta da língua virada para o céu da boca.)

fino: n.a. *þunnr*, ir. *tanae*, lat. *tenuis*, gr. *tanaós* “longo”, a.e.e. *тнѣкъ* “fino, esguio”, h.a. *tanúh*, *tánukah*. (O significado será “fino” a não ser quando for indicado.)

Essa fórmula funcional estabelece-se na posição de início de palavra ao preceder uma vogal. Sob outras condições, **t* na posição inicial terá outras representações em algumas línguas, porém em gótico e nórdico antigo, será sempre realizado como *þ*. No agrupamento **st*, em gótico e em nórdico antigo ocorre *t*, assim como ocorre *p* para **p* na **sp*; compare-se o gót. *steigan* “subir” e o gr. *steíkhein* “andar, marchar”.

Porém pelo que podemos observar a partir do terceiro EE a partir do final das palavras lat. *frater*, *mater* e *pater*, existem funções para **t* dentro da palavra que diferem daquilo que encontramos em posição inicial; além disso, observamos que o Gótico possui diferentes representantes na palavra *broþar* e *fadar*, correspondendo a um único representante nas outras línguas. As condições devem ser, portanto, diferentes nas duas palavras. Se coletarmos exemplos para esse contraste, encontraremos que aquilo que determina se encontraremos *þ* ou *d* no gótico é o lugar do acento na palavra correspondente no hindi antigo. Se o acento da palavra em hindi antigo precede imediatamente o *t*, então no gótico teremos *þ*; do contrário, *d*. Assim, *broþar* corresponde ao h.a. *bhrátā*, e *fadar* ao h.a. *pitá*. Essa relação entre as consoantes góticas e o acento índico foi descoberta pelo linguista dinamarquês Karl Verner em 1876. Já estabelecemos (p. 60) que os acentos devem ser considerados entre os elementos da expressão. A lei de Verner implica, é claro, que os acentos do hindi antigo devem ser incluídos em nossas fórmulas indo-europeias.

Verifica-se, de fato, que as relações acentuais (ou, pelo menos, os fenômenos que nos parecem ser mais diretamente relações acentuais) possuem um papel mais abrangente nas funções de EE. Foi demonstrado que a oposição no grego entre o acento circunflexo (^) e o agudo (´) das sílabas acentuadas é também encontrada ou refletida em muitas outras línguas indo-europeias (mais claramente em lituano, no qual ~, por um lado, e ´ ou ` , por outro, são distribuídos em sílabas finais assim como ^ e ´ o são em grego). A fórmula

apresentada acima (p. 64) para I.E. *-ā estabelece-se, em grego e lituano, no tipo de sílaba que leva o acento agudo quando acentuada (e isso é precisamente o que acontece na terminação do nominativo feminino). Quando *-ā estabelece-se no tipo de sílaba que leva o circunflexo em grego e em lituano, obtemos uma outra função. Em tais casos, lituano tem *o*, não *a*, correspondendo ao I.E. *-ā, assim como na terminação do genitivo feminino representado por Gr. -ās (-ēs) e em Lit. -os.

Assim, encontramos uma relação tal entre as línguas indo-europeias consideradas que um dado EE em uma língua, no contexto de outros EEs e em um dado lugar da palavra, possui uma correspondência constante para um dado EE em cada uma das outras línguas. Tal relação pode ser demonstrada para todos os EEs em todas as línguas indo-europeias. Por conseguinte, toda a extensão do sistema de elementos expressivos em cada uma das línguas mantém uma relação constante com toda a gama de EEs no sistema de cada uma das outras línguas. É justamente essa correspondência constante que chamamos de função-EE ou *função-elemento* (desde que concordemos em entender esse último termo como uma referência aos elementos da *expressão*).

O princípio por trás dessa observação pode ser generalizado. Tem-se um princípio universal: onde houver uma relação genética entre línguas, haverá função elementar entre os sistemas dessas línguas. Tal afirmação é válida para todas as famílias linguísticas, distantes ou próximas, pequenas ou grandes. O mesmo método fora utilizado nas instâncias em que foi possível estabelecer uma relação genética entre as línguas indígenas americanas. Quando estabelecemos, ou tentamos estabelecer, uma relação genética entre o indo-europeu e outras línguas, tem-se, novamente, questão de estabelecer uma função-elemento. Também uma função-elemento é utilizada para estabelecer os tipos de relações mais próximas encontradas dentro de subfamílias. Uma família de língua abrangente como o indo-europeu se compõe de subfamílias para as quais podemos estabelecer funções-elemento específicas e determinar fórmulas correspondentes. Em outras palavras, existem famílias linguísticas no interior de famílias linguísticas, ou graus de parentesco entre línguas.

Uma subfamília interna ao indo-europeu é o *itálico*, que inclui o latim e as línguas *românicas* modernas (italiano, espanhol, francês, romeno, português, etc.) Temos aqui um exemplo de função-elemento entre as línguas românicas:

*ct = it. tt φ es. ch (pronunciado como no ing. ch) φ fr. it φ rom. pt φ port. it,

como observado em:

it. <i>fatto</i>	es. <i>hecho</i>	fr. <i>fait</i>	rom. <i>fapt</i>	port. <i>feito</i>
it. <i>latte</i>	es. <i>leche</i>	fr. <i>lait</i>	rom. <i>lapte</i>	port. <i>leite</i>
it. <i>notte</i>	es. <i>noche</i>	fr. <i>nuit</i>	rom. <i>noapte</i>	port. <i>noite</i>

Assim como chamamos nossas fórmulas de indo-europeias quando estávamos comparando as línguas do indo-europeu, o mesmo fora feito pelos Romanistas ao dar às próprias fórmulas nomes, tais como romanço comum, protoromanço, ou latim vulgar. Levando em conta razões de adequação, os signos que representam cada fórmula foram escolhidos de forma a se assemelhar, o melhor possível, ao EE latino correspondente. Porém essas denominações são também arbitrárias, assim como as fórmulas indo-europeias, e há muitas instâncias nas quais será impossível alinhá-las com os EE nas palavras latinas.

Outra subfamília do indo-europeu é o *germânico* (ger.), que inclui o gótico, o nórdico antigo e as outras línguas escandinavas, e os muitos estágios do alemão e do inglês. As fórmulas dessa família são conhecidas como germânico comum ou protogermânico e, como se vê, é oportuno selecioná-las de maneira a retomar, dentre os numerosos casos (mas não em todos), os EE góticos. Dessa forma, por exemplo, o germanista não terá a necessidade de partir do I.E. **t*, mas pode estabelecer um ger. **þ* e demonstrar quais funções entre as línguas germânicas correspondem a ela sob diferentes condições. Para a condição ilustrada pelos exemplos dados acima, obtemos a seguinte função:

ger. **þ* = gót. *þ* \varnothing n.a. *þ* \varnothing escand. moderno (din., sue., nor.) *t* \varnothing i.a. *þ* \varnothing ing. moderno *th*
(desvozeado, como em *thin*) \varnothing a.a. *d*.

Ao citarmos as línguas nessa ordem e, retornando aos nossos exemplos anteriores, encontramos:

gót. *þaurusus*, etc.; n.a. *þurr* “seco”, *þerra*, *þorna* “secar”, *þorsti* “sede”, *þyrstr* “sedento”, din. *tør* “seco”, *tørre* “secar”, *tørst* “sede”; i.a. *þyrre* “seco”, *þurst* “sede”; ing. moderno *thirst*; a.a.a. *durri* “seco” *derren*, *dorrên* “secar”, *durst* “sede”; a.a. moderno *dürr* “seco”, *dorren* “ser seco”, *dörren* “secar”, *Durst* “sede”.

n.a. *þurnnr* “fino”, din. *Tynd*, ing. moderno *thin*, a.a.a. *dunni*, a.a. moderno *dünn*.

Correspondentemente, podemos estabelecer ger. **t*, a partir do ponto de vista germânico, para os exemplos apresentados acima para determinar I.E. **d*:

ger. **t* = gót. *t* \varnothing n.a. *t* \varnothing escand. moderno *t* \varnothing i.a. *t* \varnothing ing. moderno *t* \varnothing a.a. *z*,

assim como em

gót. *ga-tamjan*, n.a. *tamr*, *temja*, din. *tam*, *tamme*, i.a. *tam*, *temian*, ing. moderno *tame*, a.a.a. *zam*,
zamôn, *zemmen*, a.a. moderno *zahn*, *zähmen*.

gótt. *twai*, n.a. *tveir*, din. *to*, ing. moderno *two*, a.a.a. *zwêne*, a.a. moderno *zwei*.

gót. *taihun*, n.a. *túu*, din. *ti*, i.a. *tíen*, ing. moderno *ten*, a.a.a. *zehan*, a.a. moderno *zehn*.

As funções EE a.a. *d* \varnothing gót. n.a. *p* e a.a. *z* \varnothing gót. n.a. *t* foram reconhecidas em 1814 por Rask, no ensaio já mencionado, em conexão com outras funções EE nas mesmas línguas. Assim como ger. *f* = I.E. **p* e ger. *p* = I.E. **t* são funções particularmente marcantes e características do germânico em sua relação com as outras línguas indo-europeias, da mesma maneira a.a. *d* = ger. **p* e a.a. *z* = ger. **t* são funções particularmente marcantes e características do alto-alemão em sua relação com as outras línguas germânicas (O baixo-alemão possui *th* (e, mais tarde, *d*) e *t*, respectivamente: no antigo saxão – a forma antiga do baixo-alemão “fino” é *thunni* e “dez” é *tehan*. Essa é uma das diferenças mais imediatamente perceptíveis entre essas duas línguas da Alemanha: alto-alemão e baixo-alemão). Essas funções características do alto-alemão também incluem *t* = ger. **d* (gót. *d*), de tal forma que a oposição identificada por Verner no gót. entre *broþar* e *fadar* assume, no alto-alemão, a forma de uma oposição entre *d* e *t*: *Bruder*, *Vater*.

Naturalmente, uma única função EE em uma língua pode corresponder a dois ou mais EE diferentes em línguas relacionadas. Vimos acima (p. 64) tanto h.a. *i* quanto h.a. *a* na primeira sílaba de palavras como EEs correspondentes a *a* nas outras línguas, nas quais *a*, portanto, iguala-se a I.E. **A* em algumas palavras e a I.E. **a* em outras. Como podemos ver, uma distinção observada em uma língua pode ser completamente eliminada em outra. Assim, a distinção observada por Verner entre o gót. *broþar* e *fadar*, o a.a. moderno *Bruder* e *Vater*, bem como no i.a. em *bróþor*, *fæder*, é eliminada tanto em escandinavo (no islandês *bróðir*, *faðir*, din. *broder*, *fader*) e no inglês moderno (*brother*, *father*), uma vez que essas línguas possuem, nesse caso, um único EE correspondendo

tanto a ger. **p* como a **d*. Outro caso da mesma natureza observado em nossos exemplos é o din. *t* (escand. moderno), o qual em algumas palavras (como *tø*r, “seco” e *tynd* “fino”) corresponde a ger. **p*, I.E. **t*, e em outros (como *tam* “domar”, *to* “dois” e *ti* “dez”) a Ger. **t*, I.E. **d*. Dessa maneira, línguas geneticamente relacionadas não necessitam ter, de forma alguma, o mesmo número de EEs ou o mesmo sistema de EEs.

As subfamílias cujos exemplos vimos aqui pertencem à grande família de línguas do indo-europeu, assim como estados dentro de outro estado. Elas se estabelecem de acordo com o mesmo princípio que rege a grande família, mas independentemente dela. O germanista, ou romanista, pode trabalhar com um sistema fechado no qual as fórmulas germânicas, ou românicas, são ferramentas adequadas, sem ter que olhar além dos limites da subfamília, para os outros membros da família mais ampla. Esse fato determina uma divisão útil do trabalho entre os linguistas que se dedicam ao estudo sobre relação genética. Cada família de línguas, mesmo pequena, bem como cada parte de uma família de línguas, é como um microcosmo, organizado exatamente como as famílias maiores. Por exemplo, poderíamos considerar as línguas escandinavas separadamente do restante e estabelecer um conjunto de fórmulas comuns somente a elas. Poderíamos restringir nosso domínio ainda mais, tomando as línguas escandinavas orientais (dinamarquês e sueco) separadamente e tratá-las exatamente da mesma forma. Poderíamos ainda tomar os dialetos dinamarqueses separadamente e demonstrar como eles possuem funções-elemento mútuas que podem ser contabilizadas a partir de um sistema de fórmulas do dinamarquês comum. A partir desse ponto de vista, cada domínio basta-se em si mesmo. Ao selecionar certas funções-elemento como características peculiares de um certo grupo linguístico, uma determinada língua, ou uma dada área dialetal (como as funções consonantais que vimos serem específicas e peculiares ao alto-alemão, opondo-o a todas as outras línguas germânicas), somos levados a distinguir famílias linguísticas de diferentes graus ou de diferentes tamanhos. De modo geral é, naturalmente, mais produtivo permanecer no interior da família mais restrita possível quando se estiver tratando dos problemas internos a essa família. Determinam-se os dialetos dinamarqueses a partir do dinamarquês comum, o dinamarquês do escandinavo oriental, e este a partir do escandinavo (ou nórdico). Ademais, caso sigamos a classificação das línguas germânicas proposta nas páginas 70 e 71, o escandinavo é determinado a partir do germânico oriental, e este o será a partir do germânico, o germânico a partir do indo-europeu, e o indo-europeu a partir das fórmulas comuns de funções entre línguas pertencentes ao indo-europeu e não pertencentes ao indo-europeu. É essa gradação

contínua que determinou a escolha das línguas (p. 58) que foram usadas para exibir algumas das funções-elemento do indo-europeu. Poderíamos, é claro, comparar o dinamarquês diretamente com o latim, ou o francês com o hindi antigo, mas tais comparações trazem consigo complicações que podem ser evitadas se escolhermos, primeiramente, uma língua germânica que é mais próxima ao germânico comum que o dinamarquês, ou uma língua itálica que é mais próxima ao itálico comum que o francês. Sem um ponto de ligação intermediário, deparar-nos-íamos com as complicações de ter de dar conta de todas as instâncias nas quais o dinamarquês elimina uma distinção encontrada no germânico comum ou em gótico, ou na qual o dinamarquês introduz uma diferença que não é encontrada ali, ou, correspondentemente, na qual o francês elimina uma distinção encontrada no itálico comum ou em latim, ou introduz uma distinção que não é encontrada em nenhuma das duas línguas.

Por isso, para cada família de línguas, vasta ou reduzida, o linguista usa como suas ferramentas um conjunto ou um sistema de fórmulas comuns nas quais ele resume as funções-elemento observadas da maneira mais simples e mais econômica possível. Cada família de línguas pode ser contabilizada a partir de seu sistema, e cada período na história da pesquisa linguística possuiu o seu próprio sistema para cada família, já que simplificações posteriores são comuns. Por exemplo, no curso da última centena de anos pelo menos três sistemas claramente distintos, cada um com suas próprias nuances, foram estabelecidos para o indo-europeu. O terceiro está agora mesmo em processo de formação e discutiremos sobre ele um pouco mais a frente. Nesta seção, ativemo-nos provisoriamente a uma exposição metodológica do sistema alcançado pela linguística clássica, ou seja, o segundo de três. Em suma, esse sistema possui os seguintes aspectos:

1. *Consoantes*, isto é, entidades que não estabelecem por si mesmas uma sílaba.

Dessas, nós estabelecemos um mínimo de $5 \times 4 + 3 = 23$:

*p	*ph	*b	*bh
*t	*th	*d	*dh
*k ₁	*k ₁ h	*g ₁	*g ₁ h
*k ₂	*k ₂ h	*g ₂	*g ₂ h
*k ₃	*k ₃ h	*g ₃	*g ₃ h
*s	*p	*j	

2. *Vogais*, isto é, entidades que por si mesmas estabelecem uma sílaba. Dessas, nós estabelecemos um mínimo de $2 \times 3 = 6$:

*a *e *o
 *ā *ē *ō

3. A entidade *_A (a teoria clássica contabilizou-a entre as vogais; aqui, colocamo-la em uma classe à parte, em antecipação a futuras descobertas).
4. *Coeficientes*, isto é, entidades que por si mesmas por vezes estabelecem e por vezes não estabelecem uma sílaba. Há seis dessas, porém elas podem aparecer tanto como formadoras de sílaba como não formadoras de sílaba e, no primeiro caso, tanto como curtas quanto como longas. Quando levamos essas possibilidades em conta, temos o total de $3 \times 6 = 18$:

Formadora curta de sílaba: *i *u *y *ɨ *ɥ *ɨ̄
 Formadora longa de sílaba: *ī *ū *ȳ *ȳ̄ *ī̄ *ū̄
 Não formadora de sílaba: *i̇ *u̇ *ẏ *ɨ̇ *ɥ̇ *ɨ̇̄

A combinação de uma vogal e de um coeficiente (por exemplo *a_i *e_u *or *ā_l *ē_n *ō_m, etc.) é chamada de ditongo.

Cada um desses símbolos, então, é uma fórmula abreviada para um conjunto de funções que estão presentes sob condições mutuamente exclusivas. Obviamente, não há nada que previna o uso combinado desses símbolos quando precisarmos indicar que determinadas línguas relacionadas possuem dois ou mais EEs combinados, cada um dos quais corresponde a um símbolo particular no sistema. Na verdade, já fizemos isso no caso dos ditongos do indo-europeu, que são justamente uma combinação de tais símbolos. Fizemos o mesmo nas páginas 65 e 66 ao operarmos com I.E. *sp e *st, e também na página 68 ao operarmos com o elemento *ct do romance comum. Podemos ainda prontamente combinar mais símbolos sempre que for necessário. Não precisaremos sempre fazer isso; na verdade, é um acontecimento relativamente raro que uma palavra completa ou partes de uma palavra sejam completamente idênticas nas diferentes línguas (“idêntico” aqui deve ser entendido como “capaz de ser registrado como consistindo das mesmas entidades indo-europeias”). Se, por exemplo, olharmos para as palavras para

“lua” e “mês”, ou para as palavras que relacionamos sob as rubricas *eu*, *assassinato* e *inverno* (p. 61), vemos que não são idênticas em nenhum aspecto: elas foram formadas de acordo com as regras específicas de formação de palavra de cada língua. Apenas determinados componentes de cada palavra, determinados EEs, permanecem imutados de uma língua para a outra, e são esses que o linguista abstrai e usa como base para as suas construções. Vimos acima que as palavras listadas incluem todas o I.E. **m*, assim como observamos (p. 60, 65) que as palavras para “domar” contêm de forma idêntica um representante de **m* e um representante de **d* em todas as línguas citadas. Diversas de nossas palavras ilustrativas possuem aspectos idênticos em relação a muitos outros EEs – de outra maneira, elas não teriam servido de exemplo claro para o leitor não-especialista. Na prática, é bastante raro que possamos traçar uma palavra por toda a extensão da família indo-europeia e encontrá-la totalmente idêntica a si mesma em todas as línguas analisadas. Entretanto, temos o caso de *brother*, *mother* e *father* que podem ser traçadas como palavras inteiras, contanto que ignoremos a variante grega *phrátōr*, com sua vogal divergente na última sílaba, e *mótina* em lituano, que contém a extensão *n*. Podem, então, ser formuladas em sua totalidade como I.E. **bhrátēr*, **mātēr*, e **pAtēr*. Da mesma maneira, o romanista pode identificar cada conjunto de palavras na página 68 e representá-las por meio das fórmulas **factu*, **lacte*, **nocte*. Por outro lado, uma vez que tenhamos estabelecido todo o nosso sistema de fórmulas, poderemos por vezes perceber a utilidade de tomar uma palavra que aparece em uma dada língua – quiçá, nessa forma em particular, apenas nessa língua – e, com base em nosso conhecimento geral de funções-elemento, transpô-la para uma fórmula indo-europeia. Obviamente, nada nos impede de fazer isso. Podemos dizer, por exemplo, que enquanto todas as outras formas da palavra *irmão* representam I.E. **bhrátēr*, o grego *phrátōr* deve ser interpretado como I.E. **bhrátōr*. Dessa maneira, poderemos tomar qualquer palavra em qualquer língua indo-europeia e “traduzi-la” em uma fórmula indo-europeia.

Dissemos acima que as palavras, frequentemente, diferem em algum ponto nas diferentes línguas indo-europeias porque elas foram formadas de acordo com as regras específicas de formação de palavras de cada língua, mas isso não deve ser entendido como se as regras de formação de palavras mudassem radicalmente de uma língua para a outra. Desde que nos atenhamos às línguas que podem ser mais imediatamente descritas a partir do sistema de fórmulas do indo-europeu comum (gótico, por exemplo, mais do que o dinamarquês moderno, latim, mais do que francês), veremos, por outro lado, que a formação de palavras segue as mesmas regras em todas as línguas, mas que cada língua,

para cada palavra individual, possui uma certa liberdade de escolha em meio às muitas regras comuns. Em outras palavras, o sistema de fórmulas para os conjuntos de funções EE não é a única ferramenta de trabalho de que podemos dispor; além disso, podemos ter um sistema de regras de formação de palavras que se estabelece para todas as línguas que estão sendo comparadas diretamente. A alternância entre $*\bar{e}$ e $*\bar{o}$, que observamos acima na palavra grega que corresponde ao português *irmão*, reflete parte dessa regra. Uma importante característica presente nas regras de formação de palavras do indo-europeu consiste nas chamadas *alternâncias vocálicas*, de acordo com as quais uma certa vogal (ou, como é chamada, um certo grau de alternância) aparecerá sob determinadas condições gramaticais, dependendo também das relações acentuais e do grau de alternância encontrado em outras sílabas da palavra. Porém essas regras permitem uma certa margem de variação, de forma que algumas vezes dois graus de alternâncias apareçam na mesma palavra. Além disso, cada língua indo-europeia faz usos diferentes das mesmas regras. Não nos aprofundaremos nesse assunto aqui, mas cabe notar que a mais importante alternância vocálica no indo-europeu ocorre entre $*e$, $*o$, $*\emptyset$, $*\bar{e}$ e $*\bar{o}$. O verbo grego “voar” (veja acima na página 64) possui o grau e entre o p e o t no infinitivo presente *pétesthai*, enquanto o perfeito possui o grau o (em conjunto com uma sílaba reduplicativa prefixada *pe-*), assim em “eu voei” temos *pe-pót-ēmai*. Outra conjugação pretérita (de aspecto aorístico) do mesmo verbo possui o grau zero, ou seja, não há vogal entre o p e o t : “eu voava” é *e-pt-ómēn*.¹⁹ Quando a vogal $*e$ é a primeira parte de um ditongo, a alternância $*e : *o : \emptyset$ aparecerá certamente como uma alternância $*e\bar{i} : *o\bar{i} : *i$, ou $*e\bar{u} : *o\bar{u} : *u$, ou $*en : *on : *n$, e assim por diante. A alternância vocálica que encontramos em verbos fortes do inglês depende dessas regras de formação de palavras do indo-europeu comum. Formas como *sing* (cantar), *sang* (cantou), *sung* (cantado) contêm, respectivamente, I.E. $*em$, $*on$, e $*n$ ($*n$ aparece em Germânico como *un*).

Do que foi dito até agora, podemos concluir que *a relação genética é uma função entre línguas consistindo no fato de que cada elemento da expressão em cada uma das línguas estabelece uma função com um elemento da expressão em cada uma das outras línguas; vimos ainda que cada função-elemento em particular é condicionada pelos outros elementos da expressão que formam o ambiente do elemento da expressão analisado e pelo lugar que este ocupa na palavra.*

¹⁹ N. T.: Tendo o inglês valores aspectuais mais distantes do português, optamos por usar a versão francesa para a tradução dos termos em grego.

Isso é o que é conhecido como uma definição *operacional*: uma definição que provê um critério, um teste que decide quando uma relação genética se faz ou não presente em um determinado caso. Como pode ser observado, a definição é baseada nas funções entre línguas e nas funções entre seus elementos expressivos. Em si mesma, não diz nada sobre o tipo de função que conecta os elementos expressivos das diferentes línguas que possuem fórmulas em comum – ou sobre o tipo de função que conecta cada umas dessas línguas com o sistema das fórmulas comum. Apesar de ainda não termos examinado com atenção a natureza dessa função, deveremos, todavia, nomeá-la: chamá-la-emos de *continuação*; diremos que os elementos da expressão em cada uma das línguas *continuam* as fórmulas comuns e que o sistema de cada língua – ou, de maneira geral, que cada língua – *continua* o sistema agregado das fórmulas comuns. Posteriormente, seremos capazes de inquirir sobre o significado de tais afirmações.

Quando começamos, na página 35, a estabelecer as funções-elemento, fizemos uma observação geral de que elas só se estabelecem com restrições. As restrições nos domínios das funções-elemento devem ser levadas em conta. Elas não constituem *exceções* às funções-elemento, se por exceções entendermos violações arbitrárias e irregulares de uma regra. Elas constituem *contraexemplos*, isto é, domínios regulares estritamente definidos e estritamente delimitados nos quais as funções-elemento não se estabelecem e que devem, portanto, ser excluídos de nossos cálculos antes de estabelecermos as funções-elemento. As funções-elemento são como as leis dinamarquesas que não tem efeito sobre as Ilhas Faroé. A partir do momento que sabemos desse fato, podemos excluir as Ilhas Faroé de nossas considerações. Fora das Ilhas Faroé, tal lei mantém-se sem exceção; nas Ilhas Faroé, estabelece-se outra lei.

Os contraexemplos nunca funcionam de tal forma que certos *elementos* expressivos em uma dada língua não sejam afetados pelo princípio das funções-elemento. Já dissemos anteriormente que cada elemento expressivo em uma língua contrai funções-elemento com um ou mais elementos expressivos em outras línguas. Assim, é sempre o sistema completo dos elementos expressivos que é afetado pelas funções-elemento, e nunca encontramos um enclave, ou elementos individuais, dentro desse sistema que deveríamos excluir de nossas considerações. Não são meramente certos elementos em uma língua que possuem função com certos elementos em outra, mas todos com todos; é uma *estrutura linguística* que possui função com outra.

Os contraexemplos consistem mais no fato de que certos *signos* (certas palavras, por exemplo) em uma língua não são afetados pelas funções-elemento da mesma forma

como outros signos o são. Encontramos certos tipos de signos (incluindo-se aqui certos tipos de palavras) cujos elementos não contraem as mesmas funções-elemento que outros elementos de outros signos (outras palavras). Devemos, então, identificar tais tipos especiais de signos e excluí-los de nossas considerações quando estivermos registrando as funções-elemento. Os contraexemplos não dizem respeito aos elementos enquanto tais, mas apenas aos signos compostos pelos elementos e às funções contraídas por esses signos e seus elementos no *uso linguístico*. Antes de prosseguirmos aos contraexemplos, portanto, será necessário dizer algumas palavras sobre os conceitos de elemento e signo, sobre os diferentes tipos de funções entre elementos e entre signos que pertencem a uma mesma língua, e ainda sobre a estrutura linguística e o uso linguístico.

A estrutura linguística e o uso linguístico

Toda língua surge para nós, em primeiro lugar, como um sistema de *signos*, ou seja, um sistema de unidades da expressão que possuem um conteúdo, ou significado, vinculado a elas. Palavras são signos nesse sentido. Porém, as partes das palavras também podem ser signos: o *-s* em inglês é um signo vinculado ao genitivo (*Jack's father*) e também um signo vinculado à terceira da pessoa do singular no presente (*he write-s*). Outro exemplo é a palavra *in-ativ-ad-o*, que é um signo formado por outros quatro signos menores e diferentes entre si. Um signo pode ser formado por um elemento da expressão, *s*, cujo conteúdo vinculado a ele seja “genitivo”; ou pode ser formado – tanto no plano da expressão como no plano do conteúdo – pela combinação de dois ou mais elementos, como no signo latino *-ārum* em *bon-ārum mulierum* “das mulheres boas”, que é formado por quatro elementos da expressão – *ā*, *r*, *u*, e *m* – e três elementos do conteúdo – “genitivo”, “plural” e “feminino”.

Quando olhamos para um texto impresso ou escrito, vemos que é formado por signos e estes, por sua vez, são formados por elementos que seguem uma certa direção (a escrita latina da esquerda para a direita, a escrita hebraica da direita para a esquerda, a escrita mongol de cima para baixo – mas sempre em alguma direção). Quando escutamos um texto falado, verificamos que ele é formado por signos e que estes, por sua vez, consistem em elementos que se direcionam também temporalmente – alguns vêm antes, outros vêm depois. Os signos formam *cadeias*, e os elementos que constituem cada um desses signos formam, igualmente, cadeias. Usaremos, portanto, o termo *relação* para a função (dependência, relacionamento) entre signos ou entre elementos inclusos dentro de uma mesma cadeia: os signos, ou os elementos, são *relacionados* uns aos outros dentro da cadeia.

Se tomarmos de exemplo um signo da língua portuguesa como *pato*, podemos formar outros signos ao inserir outros elementos no lugar de cada elemento constituinte. Isso é possível ao levarmos em consideração ambos os elementos do conteúdo e os elementos da expressão, porém aqui nos atentaremos somente aos últimos, os quais são mais facilmente observáveis. Então, poderemos substituir cada elemento da expressão no signo *pato* por outros elementos da expressão para formarmos novos signos, tais como

*sato, rato, tato; peto, pito, poto; pano, pado, palo; pata, pate, pati*²⁰. Podemos então alocar o signo *pato* em uma cadeia horizontal orientada da esquerda para a direita e, então, colocar embaixo de cada um de seus elementos, em uma coluna vertical, os outros elementos que podem ser inseridos no lugar daqueles.

p a t o
s e n e
r i d i
t o l a

Chamaremos tais colunas verticais de *paradigmas*. Assim, um paradigma de elementos significará que uma determinada classe de elementos pode ser inserida no mesmo lugar de outra em uma cadeia: *p a t o* é uma cadeia; *p s r t* é um paradigma. À função entre os membros de um paradigma, daremos o nome de *correlação*. Entre *p, a, t, e o* temos uma relação; entre *p, s, r, e t* temos uma correlação.

Um fato particular ao qual devemos nos ater é que não podemos introduzir qualquer elemento em qualquer paradigma. Ao fazer isso, poderemos (fato que observaremos atentamente mais à frente) produzir signos que já foram previamente inutilizados em português moderno, como *delo* ou *quedo*,²¹ que ainda são linguisticamente possíveis segundo as regras que governam a combinação de elementos no português, porém também podemos produzir grupos de elementos, tais como *pgt*, ou *pkt*, que entrariam em conflito com tais regras. Apesar de os elementos serem possíveis, sua combinação é incorreta e o signo é impossível.

Assim, podemos construir um paradigma formado por elementos que podem ser inseridos em determinados lugares na cadeia e não em outros. A este paradigma daremos o nome de *categoria*. Dessa forma, *a, e, e i* podem pertencer a uma determinada categoria, enquanto *g e k* pertencem a outra.

Por outro lado, quando falamos sobre restrições na formação de signos, estamos fazendo uma observação muito superficial baseada em nossa primeira visada sobre a língua assim como ela surge para nós: como um sistema de signos. Se quisermos nos

²⁰ É importante salientar que nem todas as palavras apresentadas existem no léxico atual da língua portuguesa e constituem, portanto, meras possibilidades de formação.

²¹ N.T.: segundo Fonseca (1989), *delo* pode ser encontrado com dois significados: “disso” e “desde”; enquanto que *quedo* significaria “quieto”, “aquietar-se”.

aprofundar mais detalhadamente e indagar sobre as dependências nas quais essas restrições se apoiam, qual peculiaridade estrutural da língua previne um signo de ter uma formação e não outra, descobriremos, então, que existem certas regras para a construção silábica. Não pode haver em português um signo *pgt* ou *pkt* pelas seguintes razões:

1. Não pode haver uma sílaba *pgt* ou *pkt* em português. Se fosse possível, então *pgt* ou *pkt* poderiam ser então signos do português, assim como em português temos *pato*, *peta* e *pito*, ou como *delo* e *quedo*, os quais são possibilidades signícas não mais utilizadas, mas ainda assim possibilidade signícas, precisamente porque elas são formadas de acordo com as regras de formação silábica do português.

2. Não pode haver uma sílaba em português terminada em *-pgt* ou *-pkt*. Se fosse possível, então *pgt* ou *pkt* poderiam ser então signos – como *-s*, por exemplo, o qual não forma uma sílaba por si mesmo, mas pode ser usado como um signo da língua portuguesa, pois temos essa possibilidade para a terminação de sílabas em português (*gatos*, *ratos*, *chapéus*, etc.).

3. Não pode haver uma sílaba em português que comece em *pgt-* ou *pkt-*. Se assim fosse, então seria possível que *pgt* e *pkt* fossem signos do português, desde de que nesse caso eles fossem sempre usados no início da sílaba, assim como o *-s* é sempre alocado no final da sílaba.

4. Não pode haver uma sílaba em português terminada em *-pg* ou *-pk*, ou iniciada em *gt-* ou *kt-*. Se assim fosse, então, novamente, haveria a possibilidade de *pgt* ou *pkt* serem signos do português; eles apenas teriam que ser usados de uma maneira que sempre incluísse um limite entre sílabas: *apg-ta* ou *ap-gta*, etc.

Assim, observamos que as possibilidades signícas dependem diretamente das possibilidades silábicas. A estrutura da língua não é organizada de modo a ter regras especiais que governem a forma do *signo*. Um signo pode ser uma sílaba (e.g. *pá*) ou não (e.g. *-s*), e pode consistir em uma ou mais sílabas (*universo*, por exemplo, é formado por quatro). Porém, *existem* regras especiais na estrutura linguística responsáveis por organizar a forma das *sílabas*, e os signos devem, naturalmente, seguir tais parâmetros sem violá-los.

Então, o que é uma sílaba? Uma sílaba não é um signo (a palavra *categoria*, por exemplo, é formada por quatro sílabas, mas não por quatro signos). Uma sílaba pode incidentalmente coincidir com um signo, porque um signo pode ser formado por uma única sílaba (como no caso de *pá*), mas, mesmo nesse caso, a sílaba e o signo não são a mesma entidade. A terceira parte de *universo* possui a mesma sílaba que a palavra *ver*,

mas não o mesmo signo. Uma sílaba é algo diferente: é uma *unidade de elementos* que resulta do fato de que certos elementos contraem relações uns com os outros. A estrutura da sílaba depende das relações que podem ser contraídas pelos elementos – relações que são governadas por regras específicas para cada elemento ou para cada categoria de elementos. Dissemos acima que em Port. *a*, *e*, e *i* pertencem a uma categoria e *g* e *k* a outra. A categoria à qual *a*, *e* e *i* pertencem é chamada de categoria das *vogais*; já a categoria à qual *g* e *k* pertencem é chamada de categoria das *consoantes*. As vogais, cada uma delas, possui a habilidade de formar uma sílaba por si só, enquanto as consoantes não. Vogais e consoantes podem contrair relações umas com as outras dentro de uma sílaba, ou combinar-se para formar uma sílaba. Porém, existem certas restrições nas relações entre vogais e vogais e entre consoantes e consoantes dentro de uma mesma sílaba. Nem todo elemento pode se conectar com qualquer outro, mas apenas alguns com outros específicos. E as regras que governam essas relações serão tais que, por exemplo, certas consoantes podem combinar-se em determinada ordem, mas não em todas as ordens concebíveis: uma sílaba em português pode começar com *tr-* mas não com *rt-*.

Assim, cada elemento em uma língua possui seu lugar em uma determinada categoria, a qual é definida por certas possibilidades relacionais, excluindo certas outras. Todas essas categorias, em conjunto com suas definições, constituem o sistema de elementos da língua, ou o que chamamos de *estrutura linguística*. A estrutura linguística determina quais são as sílabas possíveis da língua e quais não são – e, desse modo, por sua vez, quais são signos possíveis da língua e quais não são.

Segue-se da estrutura linguística do português que podemos formar sílabas tais como *par*, *sar*, *rar*, *per*, *ser*, *rer*, *pir*, *sir*, *rir*, *pan*, *san*, *ran*, *pen*, *sem*, *ren*, *pin*, *sin*, *rin*, *pas*, *sas*, *ras*, *pes*, *ses*, *res*, *pis*, *sis* e *ris*. Essas, e outras, podem ser lidas a partir do esquema da p. 164. E, de fato, todas essas sílabas são encontradas em palavras do português. Ainda assim, uma sílaba derivada da estrutura linguística nem sempre vai ser realmente usada. A estrutura linguística permite a formação de sílabas do português iniciadas por *pn-*, mas enquanto possibilidades como *pneu* e *pneumático* existem, outras, como *pnau* e *pnoi* não são possíveis. Não se pode estabelecer regras para determinar o uso que é feito das possibilidades oferecidas pela estrutura linguística; uma possibilidade ser usada ou não é puramente uma questão do acaso.

Muito disso também é verdade em relação aos signos. Uma vez que um signo, assim como qualquer outra cadeia na língua, deve respeitar as regras de formação silábica encontradas na estrutura linguística, conclui-se que certas formações signíficas são

excluídas. Porém, conclui-se também que certas outras são permitidas. As cadeias *mar*, *sem*, *par*, citadas acima, são signos possíveis pois são também sílabas possíveis, porém pode-se verificar que sílabas possíveis, tais como *pir* ou *nar*, que não formam signos em português. Esses casos são possibilidades signícas fora de uso, e o fato de não serem usadas, enquanto outras são, é puramente acidental. Além disso, estamos livres para usá-los a qualquer momento que quisermos; podemos decidir introduzir amanhã no português as palavras *pir* ou *nar* com um significado qualquer, e isso é possível pois trata-se de sílabas possíveis – os elementos que fazem parte dessas sílabas têm a possibilidade de ocupar esses lugares na estrutura linguística.

Como dissemos anteriormente, uma língua se apresenta primeiramente como um sistema de signos. Porém, vemos agora que, de fato, a língua é primariamente algo diferente, mais apropriadamente um sistema de *elementos* designados a ocupar certos espaços definidos na cadeia, a contrair certas relações definidas em detrimento de certas outras. Esses elementos podem ser *usados*, em conformidade com as regras que os governam, para formar signos. O número de elementos e as possibilidades relacionais de cada elemento são estabelecidos definitivamente na *estrutura* linguística. A produtividade dessas possibilidades será determinada pelo *uso linguístico*.

Podemos concluir a partir disso que, se quisermos descrever uma língua, a pior maneira de executar esse procedimento será aquela que parecerá, de um ponto de vista superficial e externo, o mais natural e o único possível – ou seja: começar por uma enumeração dos signos que são utilizados nessa língua. Obviamente, se alguém quiser aprender uma língua não é suficiente conhecer a estrutura linguística – deve-se conhecer também o uso linguístico. Entretanto, o uso pressupõe a estrutura, então o trabalho será significativamente reduzido se se começa pela estrutura. O número de elementos da expressão em uma língua é notavelmente limitado: uns vinte seria o habitual; mais de cinquenta seria muito raro. O número de sílabas é comumente descrito com quatro dígitos. O número de signos, por outro lado, pode elevar-se às dezenas de milhares e é, na verdade, pela natureza da língua, ilimitado. Uma lista dos signos de uma língua – um dicionário – é necessariamente sempre incompleto, simplesmente porque novos signos são formados enquanto o dicionário está sendo compilado e muitos outros foram formados antes mesmo que o dicionário encontre seu último leitor. E, ainda assim, a língua na qual esses signos são formados é a mesma daquela de quando o dicionário começou a ser escrito. Uma língua continua a mesma enquanto a estrutura linguística for a mesma, e uma mesma língua pode ser submetida a diferentes usos linguísticos e alocada para diferentes

propósitos. O português continuará a ser português mesmo que novas palavras sejam criadas, contanto que essas novas palavras sejam formadas de acordo com as mesmas regras utilizadas para a construção de sílabas e sejam compostas pelos mesmos elementos da estrutura linguística.

Assim, é a estrutura linguística, e somente ela, que determina a identidade e a constância de uma língua. Enquanto lidarmos com a mesma estrutura linguística, podemos dizer, de forma significativa, que estamos diante da mesma língua. Sem esse critério, teríamos que nos contentar com a observação de que uma língua está em perpétuo processo de transformação e que, a cada manhã, acordaríamos com uma língua diferente daquela que tivemos no dia anterior. A cada manhã leríamos no jornal algum signo ou outro qualquer que é novo no português, ou pelo menos novo para nós. Sem esse critério, além do mais, nunca poderíamos dizer que sabemos uma língua: ninguém sabe e entende todos os signos que já foram usados em português ou todos os signos que são usados diariamente por especialistas em determinada área, ou por determinadas pessoas em certa região ou certo meio.

A estrutura linguística não somente determina a identidade de uma língua; é também o que essencialmente determina a diferença entre as línguas. Uma pessoa que sabe apenas uma língua ou algumas que possuem uma estrutura aproximada estará inclinada a acreditar que as regras combinatórias dos elementos possuem, de certo modo, uma natural inevitabilidade. Se a essa pessoa for dito que determinadas combinações não são possíveis em português, ela possivelmente estará inclinada a pensar que isso ocorre simplesmente porque a combinação não pode ser feita por lábios humanos. Na verdade, são somente seus hábitos pessoais de pronúncia que estão envolvidos nessa situação. Em relação à combinação de sons, pouco pode ser considerado impossível. Nossos órgãos fonadores são verdadeiros faz-tudos. Nós simplesmente os habituamos a produzir apenas uma pequena fração de todas as articulações que podem ser produzidas caso os tivéssemos treinado de outra forma, isto é, com propósitos além daqueles exigidos pela nossa língua nativa. *vlk* e *krk* são palavras impossíveis no português, mas perfeitamente aceitáveis em tcheco (no qual significam “lobo” e “pescoço”, respectivamente); *lgat'* e *rvat'* são inconcebíveis como palavras do português, mas são perfeitamente corretas em russo (significam “mentir” e “rasgar”); o georgiano (uma língua falada no Cáucaso) pode fornecer palavras que um falante nativo de português que, motivado por seus hábitos de fala, acreditaria ser fisicamente impossíveis – palavras como *vhsdšam* “eu como”, *mtha* “colina”, *mkbenare* “morder”, *dsqali* “água”. Para um finlandês, palavras como *skvat* ou

skvulp causariam dificuldades quase que insuperáveis, assim como *skælmisk* causaria iguais dificuldades a um lusófono, porque a língua finlandesa não permite encontros consonantais no início de sílaba e o português não os permite no fim.²²

A relação entre elementos e signos que estamos discutindo aqui é o verdadeiro segredo de todo o engenhoso mecanismo prático da língua – naturalmente, alguém poderia dizer que há um toque de gênio por trás da construção da língua. Sempre é possível formar novos signos ao se alocar os mesmos elementos de uma maneira diferente segundo as mesmas regras, e ambos, elementos e regras, são poucos e facilmente aprendidos. Dado um número limitado de elementos com suas regras de combinação, temos um inesgotável número de combinações possíveis e, portanto, de signos. O sistema de elementos é absolutamente fechado, mas o sistema de signos é produtivo; os elementos formam um conjunto fechado, enquanto os signos formam um conjunto aberto; em qualquer língua, o número de elementos é inalterável, mas o número de signos pode ser aumentado de acordo com as necessidades e os desejos da sociedade ou do indivíduo (por exemplo, no caso do poeta ou do especialista técnico), também pode haver um decréscimo, à medida que certas palavras caem em desuso e são descartadas por serem supérfluas ou indesejadas. Por ter tamanha fluidez, o sistema de signos não é meramente aplicável a certas situações, mas adaptável, sem restrições, a qualquer nova situação. Isso explica porque nenhuma língua está particularmente ligada a qualquer esfera conceitual, meio ou cultura específicos. Especialistas em línguas ameríndias têm enfatizado acertadamente que essas línguas são tão qualificadas quanto qualquer outra para expressar a cultura ocidental mesmo que, enquanto meio para expressar a cultura indígena, ainda não tenham produzido signos para muitos dos conceitos, tais como aqueles de origem técnica ou científica. Porém apenas surja a necessidade e elas terão plena capacidade de formar tais signos de maneira perfeitamente adequada. Toda língua possui, em adição aos signos que já estão em uso, uma reserva praticamente inesgotável de signos possíveis.

Consequentemente, pode-se imaginar diferentes formas de uso da língua. A mesma estrutura linguística pode ser combinada com os mais diversos usos linguísticos. Poder-se-ia tomar todas as palavras de um dicionário do português e substituí-las por outras totalmente novas e, mesmo assim, preservar intacta a estrutura linguística do

²² N.T: Procuramos refletir a mesma dificuldade de pronúncia apresentada no original na realidade de um falante de português. No original, lê-se: “A word like *skvelch*, on the other hand, would cause a Finn or a Frenchman insuperable difficulties, if only because Finnish permits no consonant clusters at the beginning of a syllable and French no consonant clusters at the end.” (ver p. 235).

português, simplesmente ao manterem-se os mesmos elementos, e apenas estes, e combiná-los em sílabas de acordo com as mesmas regras.

Porém, enquanto muitos usos linguísticos podem ser combinados a uma única estrutura linguística, o inverso não é verdadeiro. Apenas uma estrutura linguística pode ser combinada com um determinado uso linguístico. Na função entre a estrutura linguística e o uso linguístico, a estrutura é a *constante* e o uso é a *variável*. Isso é o que torna razoável dizer que é a estrutura e não o uso que determina a identidade de uma língua e a define perante outras línguas.

A estrutura linguística, então, fixa o número de elementos de forma que cada um desses elementos pode ser combinado com outros. Nada mais. Todos os outros fenômenos observáveis na língua podem variar em relação a isso e constituem, portanto, o uso. Um exemplo disso é a formação de signos e a utilização dos signos possíveis, mas outros fenômenos também são variáveis em relação à estrutura linguística e, portanto, pertencem ao uso linguístico. A representação dos elementos (veja a p. 35) não é fixada pela estrutura linguística e pode variar arbitrariamente em relação a ela. Os elementos em uma estrutura linguística podem ser representados por quaisquer meios, desde que os elementos exigidos pela estrutura se mantenham distintos. Os elementos podem, por exemplo, ser representados graficamente, sendo cada elemento representado por uma letra. Contanto que as letras sejam distintas umas das outras, elas podem assumir qualquer formato desejado. Somos livres para escolher e modificar suas formas, cores, etc. ainda assim preservar a estrutura linguística. Podemos representar a mesma estrutura linguística com um novíssimo sistema de escrita ou com um código de signos convencionalizados, e a língua, não obstante, seria a mesma. Dessa forma, é possível representar os elementos da expressão do português em alfabetos não latinos ou em código Morse e, ainda assim, termos perfeitamente a estrutura do português. Os elementos podem também ser representados foneticamente, cada elemento vinculado ao seu próprio som, não interessa qual som, desde que seja suficientemente distinto dos outros. (O ouvido humano pode ser treinado para ouvir nuances muito sutis, então o que fora dito anteriormente sobre os órgãos fonadores também se aplica aqui: o ouvido, também, é um faz-tudo, porém, o treinamento especial que ele recebeu para ouvir as nuances que estão em jogo em sua língua nativa reduz essa habilidade em relação às nuances que fazem parte de outras línguas, o que levará a afirmações de que é fisicamente impossível perceber as distinções utilizadas em qualquer outra língua que não a sua.) É igualmente imaterial o meio pelo qual os sons são produzidos – seja por meio da fonografia, gravação sonora, apito de

sinalização ou pela boca e garganta humana. Os elementos da estrutura linguística podem ser igualmente representados por outros meios, como bandeiras de diferentes formatos e cores. Essa é a maneira pela qual navios ao redor de todo o mundo comunicam-se entre si em português ou qualquer outra língua: a partir de um sistema de sinalização universal. Determinadas frotas navais possuem seus próprios códigos de bandeiras, de modo que se comunicam em suas línguas nativas. O alfabeto manual dos surdos é outra maneira utilizada para representar os elementos da expressão de uma língua. Poderíamos seguir citando vários exemplos e, limitados apenas pela nossa criatividade, seríamos capazes de construir sempre novos meios para executar a mesma estrutura linguística, novos meios para falar uma mesma língua.

Os elementos da estrutura linguística lembram-nos as entidades utilizadas em álgebra – como a , b , c , x , y , z – com as quais podemos substituir diferentes numerais em aritmética e diferentes números, com suas diferentes nomenclaturas, em raciocínios práticos. Desde que cumpramos as condições expressas, podemos representar as entidades algébricas segundo nossa vontade.

Usando uma comparação que possa talvez ser mais produtiva, podemos dizer que a língua é organizada como um jogo – como xadrez ou a um jogo de cartas. Os elementos são as peças ou as cartas. Línguas diferentes, assim como jogos diferentes, têm regras que divergem no todo ou em parte. Essas regras estabelecem como um determinado elemento, seja uma peça ou uma carta, pode ou não deve ser usado. Até certa medida, elas restringem as possibilidades combinatórias, mas o número de combinações permitidas ou de formação de signos, assim na língua como no xadrez, são enormes. A totalidade das regras, que determinam quantas peças devem estar em jogo e como cada peça pode ser combinada com outras, pode ser considerada como a estrutura do jogo; e ela difere do uso do jogo, assim como a estrutura linguística difere do uso linguístico. Um levantamento sobre o uso no jogo teria que incluir informações não só sobre o que é *permitido* fazer (essa é a estrutura do jogo), mas também como os jogadores estão acostumados a agir em determinadas situações ou de fato agiram até agora (esse é o uso do jogo) – em outras palavras, quais combinações possuem um estatuto consagrado pelo tempo em determinadas condições. Da mesma forma, um levantamento sobre o uso linguístico terá que incluir informação sobre o uso costumeiro dos signos em um dado momento no tempo em um lugar definido sob certas condições. Um levantamento do uso do jogo também deverá incluir informações sobre o material com o qual as peças são normalmente feitas, ou vêm sendo feitas até o momento, ou qual o mais apropriado, e terá que nos dizer sobre

a aparência atual de cada peça, ou tipo de peça, ou como elas se parecem até o momento (o cavaleiro do xadrez parece com a cabeça de um cavalo, e assim por diante) – e, da mesma forma, o levantamento do uso linguístico terá que conter também uma descrição do material (gráfico, fonético, etc.) com o qual os elementos são feitos, além de nos dizer também em detalhes como a parte gráfica, sonora, etc. são constituídas. Cada uso do jogo pressupõe a estrutura do jogo, mas não inversamente. Muitos usos diferentes do jogo podem ser alinhados a uma mesma estrutura. A estrutura do jogo permanece a mesma, independentemente do uso que é feito de suas regras. Essa é precisamente a razão pela qual a mesma estrutura de jogo pode ser usada indefinidamente para gerar novas combinações – novos jogos – assim como a estrutura linguística pode ser usada indefinidamente para gerar novos signos. A estrutura do jogo permanece a mesma, mesmo que as peças sejam feitas de um material diferente ou assumam uma forma divergente do usual. Nesse aspecto, a estrutura do jogo permite liberdade completa, com a única restrição de que as peças que possuam divergências em suas regras combinatórias possam ser suficientemente distinguíveis. Nesse sentido, vale lembrar que pessoas podem jogar xadrez por meio de mensagens sem utilizar nenhuma peça, mas representando os elementos e posições do jogo por meio de letras e números.²³

Ao utilizar uma língua, as pessoas começam a associar a ela, mais ou menos conscientemente, certas noções bem diferentes. Essas noções podem também ser incluídas como parte do uso linguístico; em todo caso, elas claramente não pertencem à estrutura linguística. Via de regra, são noções que estão diretamente associadas, de uma maneira ou de outra, com a forma externa dos elementos representativos – a forma dos sons, por exemplo, ou das letras. Algumas serão percebidas como belas, outras como feias, algumas são similares e outras não, algumas lembram coisas fora da língua enquanto outras não. Um exemplo dessa última situação é o chamado *simbolismo sonoro*. Palavras como *sucinto*, *esguião*, *guri* e *pino* incluem em si um elemento representado foneticamente no uso linguístico por um determinado som (na escrita, a letra *i*) e porque, de uma forma ou de outra, relaciona-se o conteúdo significativo desses signos à noção de pequenez, a impressão que fica é de que exista uma vaga conexão entre esse som e a

²³ N.T.: Decidimos atualizar o exemplo a partir de elementos mais familiares aos leitores hodiernos. No original, lê-se: “In this connexion, it may be recalled that one can play chess by telegraph, without using pieces at all but representing the elements of the game by signs in Morse code.” (ver p. 238)

noção de algo pequeno. É preciso deixar claro que essas questões de similaridade ou dissimilaridade entre som, ou entre letras, ou entre sons ou letras de um lado e coisas fora da língua de outro, são questões puramente subjetivas, que dependem inteiramente do ponto de vista a partir do qual a comparação é feita. Duas coisas nunca são similares em todos os aspectos, mas apenas em alguns; e, a depender da engenhosidade, podemos sempre encontrar alguma similaridade entre duas coisas por mais que elas sejam diferentes. Em minha mesa tenho um objeto que, em relação à forma, assemelha-se a um cachorro; em relação ao material, assemelha-se mais a um amontoado de limpadores de cachimbos; em relação à cor, assemelha-se a neve suja; e assim por diante. Da mesma forma, com imaginação suficiente, pode-se encontrar semelhança entre qualquer dupla de sons da fala ou qualquer dupla de significados.

Todas essas noções ligadas à aparência externa dos elementos representativos podem, é claro, ser ligadas às suas combinações usuais permitidas pela estrutura linguística – aos signos, por exemplo, às combinações de signos, aos turnos das falas e assim por diante. Porque estão em estrita conexão com a própria matéria da língua, ou porque os usuários da língua em um determinado momento e em um determinado lugar por uma razão qualquer assim decidem, muitas dessas noções podem se tornar coletivas. Assim, a maioria dos membros de uma coletividade linguística crerá de uma maneira mais ou menos clara nessas noções, ou poderão ser facilmente persuadidos a fazê-lo. Nomeamos a totalidade dessas noções de mesma natureza de *familiaridade linguística* ou um *Sprachgefühl*. Obviamente, trata-se de um fenômeno psicológico, cujo estudo requer a colaboração de psicólogos e linguistas; por ora, é, em grande parte, terreno inexplorado.

Formação dos signos

Levando-se em consideração que a comunidade linguística é livre para adicionar ou descartar signos conforme sua vontade, é óbvio que essa familiaridade com a língua, especialmente quando se refere a um sentimento coletivo, será um fator decisivo na formação dos signos. Os contraexemplos às funções-elemento, sobre os quais discorreremos mais a frente, ocorrem justamente porque os signos podem ser traduzidos, ou criados, com base em um determinado *Sprachgefühl*. Dessa maneira, como vimos, tal familiaridade com uma língua é um território inexplorado e, por sua própria natureza, um tanto intangível, não possuímos o que seria necessário para produzir uma delimitação precisa desses contraexemplos. Se, no entanto, podemos fazê-lo, é porque podemos nos contentar, em grande medida, em substituir esse sentimento linguístico pelos fatores objetivos (fonéticos, linguísticos etc.) aos quais ele está ligado.

Agora examinaremos os tipos mais importantes de formação de signos e demonstraremos como eles produzem contraexemplos às funções-elemento.

- I. A TRANSFORMAÇÃO
FUNCIONAL DOS SIGNOS
- I. 1. A influência dos elementos

Dentro de uma cadeia de elementos (em uma palavra, por exemplo), um elemento pode ocasionar a presença em um outro lugar dessa mesma cadeia de um elemento diferente daquele que seria esperado. Por exemplo, correspondendo à preposição que aparece no got. *fair*, ir. *air-*, gr. *pér, péri, perí*, lit. *per-*, a.e.e. *prě-*, e i.a. *pári* (que possuem um significado aproximado nessas línguas de “em torno, através, por meio de”), as funções-elemento do indo-europeu nos levariam a esperar que palavra do latim *per* surgisse com um *r* de acordo com a função apresentada na p. 41. E, naturalmente, a palavra realmente é *per* em latim. Agora, podemos combinar essa palavra com a palavra *ager* “campo, território, território campestre” para formar o verbo *per-agrāre* “atravessar um país”, isto é, “viajar através, caminhar por” e o advérbio *per-egrī, per-egrē* “fora do país” (o qual possui um *e* no lugar de um *a* sob condições que deixaremos de lado por enquanto), do qual deriva *per-egrīnus* “estrangeiro, estranho”. Tudo isso ocorre de acordo com as regras das funções-elemento. Porém, essa palavra pode surgir em latim também como *pelegrīnus*, forma esta que conhecemos por meio de manuscritos. E todas as

diversas formas que encontramos em cada língua latina (it. *pellegrino*, esp. *pelegrinom*, fr. *pélerin*, port. peregrino) pressupõem a fórmula do latim vulgar **pelegrinu*. Essa provavelmente também é a forma latina que fora introduzida no germânico, como em a.a.a. *piligrîm*, em din. *pil(e)grim* e em ing. *pilgrim*. Entretanto, essa forma com *l* não está de acordo com às funções-elemento. Isso é causado pela influência de um elemento na palavra, o *r* na sílaba *-grī-*, que causou essa aparição inesperado do *l* no lugar do *r* na sílaba *-re-*. Dos exemplos citados, apenas o português não segue essa lógica.

A razão para a existência desse contraexemplo é, obviamente, para evitar que o mesmo elemento, *r*, apareça em dois lugares na mesma palavra. O contraexemplo possui uma causa psicológica, porque, normalmente, sob certas circunstâncias, é mais difícil manter dois elementos do mesmo tipo do que dois diferentes. Como estamos lidando com um aspecto psicológico, nunca poderemos prever quando isso ocorrerá ou não. Mas foi provado que é possível estabelecer regras gerais pelas quais podemos prever (1) qual estrutura em particular caracteriza as palavras que podem ser substituídas por tais transformações, de maneira que podemos desconsiderá-las quando analisamos as funções-elemento, e (2) qual elemento em uma palavra de uma dada estrutura será afetado por uma mudança caso essa mudança venha a acontecer. Antes de demonstrarmos isso, podemos estabelecer que, em geral, o elemento substitutivo introduzido quando um signo é assim transformado sempre será um elemento que já fora admitido pela estrutura linguística em questão. A transformação nunca envolve um acréscimo ou um decréscimo no inventário de elementos da língua. Além disso, ela ocorre de acordo com as regras para formação de signos que podem ser deduzidas das regras da estrutura linguística que rege o uso dos elementos. Dentro desses parâmetros, entretanto, existe uma certa liberdade em relação à *escolha* do elemento substitutivo (desde que a estrutura linguística permita a escolha de duas ou mais possibilidades nesse lugar determinado da sílaba). Os falantes escolhem o elemento que parece para eles, com base no sentimento de língua que possuem, o mais próximo daquele que será substituído. Entretanto, sempre é possível apontar razões físicas (ou, mais comumente, psicológicas) para explicar o porquê um som soará aos falantes da língua como sendo especificamente o mais próximo de outro som, de forma que, dado um conhecimento geral sobre o sistema fonético da língua, podemos prever com uma certa margem de certeza (em muitos casos, com absoluta certeza) qual elemento será escolhido como substituto. Caso façamos tal cálculo preliminar sobre o *r* tomando como base uma análise fonética do sistema de sons do latim (o que não faremos aqui), chegaremos à conclusão óbvia de que o único substituto possível é o *l*.

Três fatores serão decisivos em relação a qual elemento em uma palavra de uma determinada estrutura será afetado pela mudança, caso ela ocorra:

1. se os elementos envolvidos entram em uma sílaba tônica ou átona;
2. se eles são encontrados no *início*, *centro* ou campo *final* da sílaba (em uma palavra como *charco*, temos *c* e *h* no campo inicial, *a* no campo central e *r* está no campo final da sílaba; em uma sílaba como *ar*, não temos campo inicial, *a* está no campo central e *r* está no campo final; na sílaba *pá*, vemos que *p* está no campo inicial e *a* está no campo central, enquanto que o campo final não se faz presente; na sílaba *a*, ambos os campos inicial e final não estão presentes);
3. se eles são estão *sozinhos* no campo (como *r* em *ar*, *p* em *pá* ou *c* e *r* em *carro*) ou em grupo (como *c* e *h* em *charco*) e se estão *cobertos* ou *descobertos* (um elemento coberto é aquele que está no campo inicial que surge imediatamente após um elemento que está no campo final de uma sílaba: o segundo *c* em *charco* é coberto, enquanto que o *c* em *ecologia* é descoberto).

Com a ajuda dessas definições, poderemos estabelecer leis gerais que estabeleceram qual elemento em uma palavra pode ser submetido à influência de outro elemento. A seguir, examinaremos alguns exemplos de tais regras gerais.

I. 1 a. *Dissimilação*

Na dissimilação, os falantes evitam repetir os mesmos movimentos fisiológicos para representar dois elementos. A palavra *pelegrīnus* é um exemplo disso: a troca do *r* pelo *l* elimina a repetição do movimento da língua que é necessário para articular *r*. Outro exemplo vem do lat. *tenebrae* “escuridão”, que corresponde ao i.a. *tāmīsrā*. Das funções-elemento (cf. p. 38) deveríamos esperar um *m* em latim correspondendo ao *m* do hindi antigo. O *n* que observamos em latim foi introduzido pelo processo de dissimilação, assim eliminando a repetição dos movimentos dos lábios que são usados para a pronúncia tanto de *m* como de *b*. Desse último exemplo, vemos que a dissimilação pode acontecer até mesmo se dois elementos não são completamente idênticos, contanto que elemento possuam em comum alguns movimentos específicos nos órgãos da fala. Entretanto, na maioria dos casos de dissimilação há uma similaridade completa, de forma que a maioria do vocabulário que é afetado pela dissimilação – portanto deve ser desconsiderado quando tratamos das funções-elemento – consiste de palavras nas quais o mesmo elemento surge duas vezes, com um ou mais elementos entre eles.

O fato de que é o primeiro *r* em *peregrīnus*, e não o segundo, que é afetado pela dissimilação ocorre devido a lei geral de que um elemento em um grupo de uma sílaba tônica dissimila um elemento que está sozinho entre duas vogais – nunca o inverso. Assim, é inconcebível que a forma *peleglīnus* surja a partir da dissimilação; portanto se *peregrīnus* sofre dissimilação, *pelegrīnus* é o único resultado possível. Por resultado dessa mesma regra geral, *cérébral* “cerebral” é dissimilado no francês popular em *célébral*, *contrario* “contrário”, em galício (falado no noroeste da Espanha), torna-se *contralio* e *imperatrice(m)* “imperatriz” torna-se *empānatriz* em português de Portugal.

O fato de que em um caso como *tembrae* o *m*, e não o *b*, é o elemento que poderia ser afetado (e na verdade o é) pela dissimilação pode ser explicado pela lei geral que, de dois elementos possuidores da mesma definição de acordo com o critério da página 66, é sempre o primeiro, nunca o segundo, que é afetado por esse processo. Em *tembrae*, *m* e *b* possuem a mesma definição: ambos estão em sílabas átonas (o acento tônico está na primeira sílaba da palavra), ambos estão no campo inicial de suas respectivas sílabas e ambos estão descobertos. A sílaba que contém *m* vem antes daquela que contém *b*, portanto, será aquele elemento que passará por dissimilação, não este. Seguindo a mesma lei, em fr. *militaire* é dissimilado na fala coloquial em *mélitaire*: ambos *i*'s estão em sílabas átonas e em campos centrais – conseqüentemente é o primeiro *i* que passará por dissimilação, não o segundo.

Se no fr. *barbier*, derivado de *barbe* “barba”, pode aparecer em alemão como *Balbier* e em dinamarquês como *balbér*, mas não pode surgir como *barbiel* ou *barbél*, isso é explicado pela lei geral de que um elemento em um campo final de uma sílaba tônica dissimila um elemento que está no campo final de uma sílaba átona, e não o inverso. Isso explica o porquê de observarmos em latim *arbor* “árvore” – ou em francês *arbre* – e vemos em espanhol *árbol*, também podemos citar o caso de *marmor* “mármore” em latim que surge em espanhol como *mármol* e em alemão e dinamarquês antigo como Marmel.

Outra lei geral estabelece que um elemento coberto dissimila um elemento descoberto, nunca o inverso. Isso explica a mudança que ocorre do latim *quīnque* “cinco” para *cinque*, o qual forma a base para o fr. *cinq* e o it. *cinque* etc. A palavra *tatufolo* em italiano foi incorporada no alemão como *Tartuffel*, mas, de acordo com a regra que citamos acima, fora mudada tanto em alemão como em dinamarquês para *Kartoffel*, “batata”. Dissimilações como *quince* ou *tarkoffel* são inconcebíveis.

I. 1 b. *Metátese*

A metátese (no sentido restrito que usaremos aqui) é a transferência de um elemento de uma sílaba para outra. É observada em duas formas:

1. Transferência de apenas um elemento. Esse caso sempre consiste na transferência de um elemento de uma sílaba átona para uma sílaba tônica – nunca o inverso. Por esse processo, a língua italiana adquiriu a palavra *cocodrillo* no lugar de *crocodīlus*, como surge em latim. A forma que contém metátese evidentemente existiu no latim tardio e foi levado ao alto-alemão médio no qual temos *kokodrille*.

2. Intercâmbio de dois ou mais elementos. Nos casos mais simples e mais comuns, dois elementos em sílabas vizinhas são intercambiados. Isso sempre ocorre de modo que os elementos que não aparecem na ordem de expiração são transpostos para que assim o façam. A ordem de expiração diz respeito à ordem dos movimentos dos órgãos da fala: do interior para o exterior – da garganta para os lábios. Vimos acima (p. 41) a palavra latina *specere*, uma palavra também observada no hindu antigo como *spásati*. Essas formas estabelecem a fórmula indo-europeia da raiz como **spek₁-*. Mas em grego, a palavra que significa “olhar, ver, examinar” é *sképtesthai* (da qual é derivada a bem conhecida palavra *sképsis*, que significa “examinar, especular” e *skeptikós* “aquele que especula ou reflete” que deu origem ao inglês *skeptic*). Isso é resultado da metátese pela qual a sequência *p-k* passou para se tornar *k-p*, porque o som de *k* é formado em uma parte anterior àquela onde o som de *p* é formado: *p* é um som bilabial, enquanto que *k* é produzido por oclusão entre a língua e o palato.

Outros tipos de metáteses são governados por outras leis que não trataremos aqui.

I. 1 c. *Haplologia*

Esse termo refere-se a omissão de uma sílaba causada pela semelhança que esta possui com a sílaba vizinha: *tragicômico* em vez de *trágico-cômico*, *mineralogia* em vez de *mineralo-logy*. A maioria das leis gerais que controlam a influência dos elementos que conhecemos foi descoberta pelo estudioso francês Maurice Grammont. Algumas dessas leis podem exigir um exame mais minucioso, mas não há dúvida de que leis gerais podem ser estabelecidas. Exceções aparentes podem ser explicadas pelo fato de que uma palavra sofrerá sucessivas transformações, de forma que uma determinada forma não será passível de explicação meramente a partir da aplicação mecânica de uma única lei. Além

do mais, Grammont previu a possibilidade de que a aplicação de uma lei de forma descuidada poderia levar a resultados que divergem da estrutura linguística ao produzir tipos de sílabas inadmissíveis. Em tal caso, ele calculou que a tendência à dissimilação ou à metátese, se suficientemente forte, terá seu efeito sob uma série de regras opostas àquelas que discutimos acima; mas, mesmo nesse caso, não haverá suspensão de nenhuma lei, uma vez que tais casos podem ser claramente circunscritos e definidos.

I. 2. Formação analógica

Por formação analógica, referimo-nos à influência dos signos. Quando existe um paradigma de signos (de palavras, por exemplo), um membro do paradigma pode ocasionar o surgimento, em outro lugar do paradigma, de um signo diferente daquele que seria esperado. Existem três tipos principais de formação analógica.

I. 2 a. Nivelamento

O nivelamento ocorre quando dois membros de um paradigma são trazidos para perto um do outro. No nominativo singular, o pronome pessoal da segunda pessoa do indo-europeu aparece como **tu* e **tū*, como é atestado pelo gót. *þu*, ir. *tú*, lit. *tù*, a.e.e. *ty* etc.

Os outros casos são formados a partir de um radical formado por **t* ou **tʰ* seguido por uma vogal. No acusativo, essa vogal é **e* ou **ē*, assim teremos nesse modo **te* ou **tē* – encontrados, por exemplo, em lat. *tē* e em n.a. *þi-k* p. 47 (temos uma correspondência entre ger. *-k* e gr. *-ge*, uma partícula que pode ser adicionada a pronomes; veja o gót. *mi-k*, p. 36) – e as formas **tʰe* ou **tʰē* – encontradas na base de formas como as vistas em hit. *twel* “te, t, teu”, *twetaz* “de ti”, ou i.a. *tvam* “tu”. Em grego, alguns dialetos possuem a forma acusativa **te* – como vemos no dórico *té* – e outros possuem a forma **tʰe* – como no ático-iônico *sé* (onde *s-* = IE **tʰ-*); outros casos ainda com a mesma palavra são formados com **t* e **tʰ-* utilizando-se da mesma distribuição dialetal (dativo: dor. *toí*, ático-iônico *soí* i.a. *tve*). Tudo o que foi descrito está de acordo com as funções-elemento. Os paradigmas, portanto, estão também de acordo com o que previmos:

n.a.

lat.

dor.

nominativo	<i>bú</i>	<i>tū</i>	<i>tú</i>
acusativo	<i>þi-k</i>	<i>tē</i>	<i>té</i>

Por outro lado, em gótico o acusativo não é *þi-k*, como vemos no nórdico antigo e como esperaríamos que fosse a partir da forma indo-europeia **te-ge*, mas sim *þu-k*. Um processo de nivelamento ocorreu e com isso aproximou na forma os dois membros do paradigma e, dessa maneira, produzindo um contraexemplo para as funções-elemento: um *u* ger. não pode ser um **e* (ou **ē*) em i.e. Em hitita, encontramos algo ainda mais estranho: o nominativo é *zik* (=i.e. **te-ge*) e o acusativo é *tuk* (=I.E. **tu-ge*). Esse paradigma é resultado de dois nivelamentos contraditórios, cada um se esforçando para aproximar na forma os dois membros do paradigma, o que, paradoxalmente, resulta em um efeito final que não é melhor em relação ao inicial. O grego também possui um nivelamento nesse paradigma, mas que surge de outra maneira: o nominativo do dialeto ático-iônico não é *tú* como em dórico, mas *sú*, o que demonstra a influência analógica desses casos nos quais as palavras são iniciadas em *s-* (=i.e. **tū*).

Outras instâncias de nivelamento que produzem contraexemplos às funções-elemento podem ser observadas nos numerais. Em indo-europeu, “quatro” e “cinco” são **k₃etūōres* (veja o lat. *quattuor*, lit. *keturì*, i.a. *catvārah*) e **penk₃e* (lit. *penkì*, i.a. *páñca*), respectivamente. Ambas as palavras se iniciam com duas consoantes bem diferentes, para as quais as funções-elemento de início de palavra assinalarão correspondências bem distintas em germânico. A função **k₃-* do I.E. aparece em germânico como *hw-* como é observado no pronome interrogativo: got. *hwas*, genitivo *hwis*; lat. *quis*, feminino *quae*; lit. *kàs*; i.a. *kaḥ* etc. A função **p-* do indo-europeu surge em germânico como *f-*, como vemos em *father* e em exemplos similares. No entanto, não apenas o numeral *five*, mas também o numeral *four* começam com um *f-* em germânico: got. *fidwor* “quatro”, *fimf* “cinco” – o *f* final em *fimf* corresponde à função **k₃* do I.E. e está de acordo com uma função-elemento: **k₃* possui uma correspondência representada por *f* se é precedido pelas funções **p* ou **u* na mesma palavra – cf. got. *wulfs* “lobo”, lit. *vil̃kas*, a.e.e. *vl̃kb̃*, i.a. *ṽrkah̃*, I.E. **ũl̃k₃os*. Podemos também assinalar que o primeiro *qu* no lat. *quīnque* corresponde em I.E. à **p* em acordo também com uma função-elemento: **p* possui uma correspondência em latim, *qu* ou *c*, a depender da vogal que a sucede, se há **k₃* na mesma palavra – cf. *coquere* “cozinhar”, a.e.c. *pek₃*, i.a. *pácati*). A razão para observarmos *f-* no got. *fidwor* e no ing. *four*, em vez de *hw-* em got. e *wh-* em ing., é o nivelamento analógico que aproximou na forma dois membros do paradigma dos numerais.

Em geral, nivelamentos são comuns no paradigma dos numerais. O numeral *nine* “nove” começa com *n- (cf. por exemplo, lat. *novem*, i.a. *nàva*), enquanto que o numeral *ten* “dez” inicia-se com *d- (veja a p. 41), mas nas línguas bálticas e eslávicas ambos os numerais começam com *d- (lit. *devynì*, *dėšimt*, a.e.c. *devętb*, *desętb*): o *d- em “nine” foi introduzido por nivelamento. De acordo com as funções-elemento, deveríamos esperar que o lat. *quīnque* (com a dissimilação apresentada na p. 67) terminasse em -e em espanhol. Na verdade, temos um -o (*cinco*) sob a influência analógica de *cuatro* “quatro”. Similarmente, observamos no italiano um -i no final de *dieci* “dez” (lat. *decem*) derivado do nivelamento com *venti* “vinte” (lat. *uīgintī*).

Tais paradigmas como os que vimos demonstrando até o momento – o paradigma casual de um pronome e o paradigma dos numeras – articulam-se na própria estrutura linguística e constituem absolutamente uma série fechada com um número limitado de membros (mesmo que os números formem uma série infinita, a série de *palavras* que expressam esses números é finita e, é claro, consideravelmente curta, com a maioria dos números sendo designada pela combinação de duas ou mais palavras). Porém, no uso linguístico, encontramos muitos paradigmas que possuem um número ilimitado de membros porque, como vimos, o sistema de signos não é fechado, mas produtivo. Dessa maneira, a familiaridade linguística – Sprachgefühl – será incerto na sua seleção arbitrária do paradigma ao qual um determinado signo será atribuído e essa escolha será feita com base em associações que serão, na maioria das vezes, muito vagas. A palavra para “olho” no indo-europeu começa com *o- (cf., por exemplo, lat. *oculus*, a.e.c. *oko*) e a palavra para “orelha” começa com *au- (cf. lat. *auris*, lit. *ausis*); se em germânico ambas as palavras começam com *au- (got. *augo*, *auso*, dan. *øje*, *øre*), isso é ocasionado por um nivelamento que, levando em consideração a familiaridade com a língua, tende a classificar as duas palavras no mesmo paradigma. Se a palavra alemã *Sauerkraut* foi levada para o francês como *choucroute* é porque o Sprachgefühl a classificou como pertencente ao mesmo paradigma de *chou* “repolho” baseado na similaridade de significado e, dessa maneira, aproximou na forma os dois membros do paradigma. Da mesma forma, em francês *mousqueton* “carabina”, um derivado de *mousquet* “mosquete”, foi levado para o alemão como *muskedonner* sob a influência de *Donner* “trovão” e em norueguês e sueco como *muskedunder* sob a influência da palavra *dunder* “barulho, alvoroço”. Esse tipo de nivelamento, o qual depende de associações vagas associadas a essa familiaridade com a língua e a ligação arbitrária de dois dignos em um único paradigma, é chamada de *etimologia popular*.

Os nivelamentos são as únicas formações analógicas que comumente produzem verdadeiros contraexemplos às funções-elemento. Por essa razão, consideraremos apenas brevemente os outros dois tipos de formações analógicas.

I. 2 b. Contaminação

Por contaminação, queremos nos referir a junção de dois membros de um paradigma em um único, como acontece entre *descrever* e *viver*, que formam *descreviver*, ou *tristeza* e *circundante*, que formam *circuntristeza*, exemplos retirados de Guimarães Rosa; ou ainda o *abensonhar* de Mia Couto, formado por *abençoar* e *sonhar*.

I. 2 c. Formação proporcional

Na formação proporcional, um signo é transferido a um paradigma diferente daquele ao qual ele pertencia originalmente. Em português temos o exemplo da derivação imprópria que, a depender do contexto, altera a classe da palavra. Nesse caso, por exemplo, um pronome indefinido, *nada*, pode ser transferido para o paradigma dos substantivos “um nada me atingiu”. As diferentes declinações e gêneros gramaticais também constituem diferentes paradigmas. Em alemão, a palavra *Nacht* “noite” pertence ao paradigma dos substantivos femininos que recebem o umlaut no plural (*Nächte*), enquanto que *Tag* “dia” pertence ao paradigma formado pelos substantivos masculinos que possuem uma “forte” declinação (*Tage* no genitivo) e não recebem umlaut no plural (*Tage*). O genitivo de *Nacht* na forma definida é *der Nacht*, mas mesmo assim “de noite” pode ser *des Nachts* assim como “de dia” pode ser *des Tages*: uma formação proporcional trouxe *Nacht* ao mesmo paradigma de *Tag*.

Chamamos esses processos de formação proporcional porque eles pressupõem a construção de uma proporção ou uma equação do tipo $a : b = c : x$, com a solução $x = b'$; por exemplo, $hinder : hindered = help : x$, então $x = helped$. A contaminação, por outro lado, segue a seguinte fórmula: $a \times b = c$, então, $abençoar \times sonhar = abensonhar$.

Levando em consideração que o sistema de signos está sujeito à arbitrariedade do uso linguístico e que a familiaridade linguística – Sprachgefühl – é imprevisível, fica claro que não se pode estabelecer leis gerais nessa área como é possível para a influência dos elementos. Mas, porventura, isso é causado apenas porque o conteúdo linguístico ainda não foi adequadamente analisado a partir de um ponto de vista psicológico. Se a influência

dos elementos já foi tão minuciosamente explorada, é precisamente porque somos capazes de analisar os sons da língua e porque nós conhecemos um pouco sobre a psicologia do som. Eventualmente no futuro poderemos ter bases mais sólidas do que as que temos atualmente para o estudo das formações analógicas. Enquanto isso, nossa experiência com o estudo genérico de uma enorme quantidade de material linguístico demonstrou que as formações analógicas de forma alguma nos impedem de entender as funções-elemento e, em geral, não causam nenhuma dificuldade prática em nossas investigações. Temos apenas que manter um olhar atento sobre o fato de que o princípio da formação analógica provê uma razão para deixar de lado certos contraexemplos com os quais talvez lidaremos.

I. 3. Abreviação dos signos

Um signo pode sofrer abreviação quando é transferido de certas formas para uma categoria diferente da qual ele originalmente pertencia. As ocorrências mais relevantes são as seguintes:

1. Quando uma palavra composta se torna simples. Assim, palavras originalmente compostas como “vossa mercê” e “em boa hora” foram simplificadas em português e tornaram-se atualmente *você* e *embora*.

2. Quando uma palavra derivada se torna primitiva. Assim, em ale. *fressen* “comer (animais), amontoar” substituiu *ver-essen* (got. *fra-itan*), e o ale. *gleich* “igual, parecido” substituiu *ge-leich* (got. *ga-leiks*; cf. sem o prefixo derivacional, din. *lig*, ing. *like* e lit. *lýgus* “reto, suave, parecido”).

3. Quando um signo é transferido para outra categoria gramatical e é encurtada como resultado do processo. Assim, um substantivo transmutado em uma preposição é normalmente abreviado. O substantivo lat. *casa* “casa” surge em francês antigo como *chiese*, de acordo com as funções-elemento envolvidas. Mas a frase *in casa* seguida por um genitivo, significando “em fulano de tal”, fora transferida para a categoria das preposições e assumiu o significado de “na (casa de fulano de tal)”. Como resultado disso, a vogal final caiu, resultando em f.a. *en chies* e, em francês moderno, *chez* (sem *em*). A preposição dinamarquesa correspondente, *hos*, carrega a mesma relação com o substantivo *hus* “casa”; cf. também din. *til*, ing. *till*, e ale. *Ziel* “objetivo, alvo, prazo”.

Outro caso do mesmo tipo é quando um verbo é transferido para a categoria dos verbos auxiliares – um processo de gramaticalização. O *b* do latim quando entre duas

vogais aparece como *v* em português: lat. *faba*, port. *fava*; lat. *dēbēre*, port. *dever*; lat. *habēre*, port. *haver*. Porém, quando comparamos o lat. *habeō* “eu tenho”, verificamos que em português temos *hei* com um morfema zero no lugar do *v* que seria previsto para essa posição. A razão para isso é que o verbo se tornou um auxiliar e foi, conseqüentemente, abreviado. Este processo estende-se ainda ao ponto de levar o verbo auxiliar à categoria dos sufixos, aprofundando ainda mais sua gramaticalização: *hei de ir = irei*; *hei de voltar = voltarei*. Observamos esse mesmo processo em francês ao compararmos o lat. *habeō* com o fr. *ai*: percebemos a ausência do morfema *v* causado pela transferência do verbo para a categoria dos auxiliares, o que causa sua abreviação. Dessa forma, devemos esperar encontrar em francês *j'ai* como um verbo auxiliar (como em *j'ai mange* “eu como”) e outra forma, contendo *v*, usada como verbo pleno. O fato de observarmos a forma *j'ai* nessa função (*j'ai um ami* “eu tenho um amigo”) também é por causa de uma formação analógica generalizante.

As categorias que consideramos até o momento – compostas, derivadas, substantivos e preposições, verbo pleno e verbo auxiliar – são dadas na própria estrutura linguística. Porém, o uso também pode estabelecer categorias. Entretanto, tal formação de categorias, é facilmente delimitada: uma categoria pode ser estabelecida quando formada por signos que estão em um uso frequente e especial e que, conseqüentemente, são abreviadas.

4. Um signo é abreviado quando seu uso é frequente. Podemos iniciar por alguns casos limítrofes que podem envolver a transferência de uma dada categoria na estrutura linguística. Em latim, o imperativo de verbos com radical consonantal (terceira conjugação) termina em *-e* (=I.E. **-e*): *scrīb-e* “escreva!” do verbo *scrībere*. Mas os cinco verbos *dīcere* “dizer”, *dūcere* “liderar”, *facere* “fazer”, *ferre* “carregar”, and *emere* “comprar” possuem as seguintes formas imperativas: *dīc*, *dūc*, *fac*, *fer*, e *em*, todas sem *-e*. A forma *em* possui o significado especial de “ali! veja!” enquanto que “*compra!*” é *em-e*. Esse verbo é encontrado em outras línguas indo-europeias com o significado de “tomar”: lit. *iñti*, a.e.c. *ęti*. A partir do que foi dito, podemos dizer que esses verbos imperativos foram transferidos para a categoria das interjeições, a qual certamente está inclusa na estrutura da língua. Observamos a mesma situação quando uma palavra-resposta como *sim* e *não* surge, como costuma acontecer, em uma forma que entra em conflito com as funções-elemento. Palavras-resposta também constituem uma categoria, a qual é definida por sua relação especial com os meios em que surge e a qual também é encontrada na estrutura linguística. A essa categoria também pertencem os pronomes de

tratamento, nos quais observamos uma abreviação muito acentuada: port. *senhor* torna-se *seu*; ing. *madam* torna-se *mam* ou ainda somente *m*, especialmente quando combinado com uma palavra-resposta (*yes 'm*). Do mesmo modo, em rus. *gosudár'*, primeiramente abreviado como *súdar'*, foi posteriormente abreviado a um simples *s* quando combinado a uma palavra-resposta (*da-s* 'sim, senhor'). Nesse contexto, comumente observamos o surgimento de pares, um dos quais retém seu significado lexical completo enquanto que o outro é reduzido a uma forma mais ou menos vazia de polidez. Palavras que significam “você” também se revelam como abreviações de formas mais longas e complexas: em português, *você* é a forma comprimida de *vossa mercê*; já em lituano observamos que *Támsta* é a forma abreviada de *tàvo mýlista* “tua graça”.

Mas mesmo que exista aqui uma clara conexão com as categorias que estão ancoradas na estrutura linguística, devemos ainda reconhecer no uso linguístico a existência de uma categoria de signos que são usados com uma certa frequência e, por isso, abreviados. Consequentemente, não só ing. *automobile vehicle*, fr. *voiture automobile* são abreviados para *automobile*, mas este é ainda mais abreviado – em francês, inglês, alemão e diversas outras línguas para *auto* e em dinamarquês para *bil*. Em português brasileiro, *motocicleta* torna-se *moto* pelo mesmo processo. De maneira análoga, também vemos em dinamarquês que *flyvemaskine* “máquina voadora” é encurtado e torna-se *fly*. Abreviações desse tipo são comumente encontradas em jargões profissionais porque são usados com uma frequência elevada nessas áreas (mesmo que soem estranhas ou sejam completamente desconhecidas por pessoas de fora). O jargão médico é um exemplo disso em português: *UTI* para *unidade de terapia intensiva*, *RCP* para *ressuscitação cardiopulmonar* e *seps* para *septicemia*. Ainda em português, temos *prof* para *professor*. Via de regra, para esses casos, a versão mais curta também coexiste com a longa e não possui um significado diferente; não se trata de um novo signo, mas sim de uma variante desse signo inicial, na mesma linha das abreviações na língua escrita. Além disso, tais formações, diferentemente de muitas outras, são produzidas conscientemente por uma intervenção voluntária na língua. Algumas comunidades linguísticas estão mais inclinadas a introduzi-las que outras e isso normalmente depende da estrutural geral dos signos na língua. O inglês, por exemplo, já possui tantas palavras monossilábicas que parece natural introduzir ainda mais novos signos do mesmo tipo e, por isso, a língua favorece esse tipo de abreviação. O dinamarquês é consideravelmente mais conservador a esse respeito, fato que é indubitavelmente ligado a estrutura da língua, de acordo com a qual as abreviações envolvem uma respectiva mudança na pronúncia. A

despeito de sua praticidade, *fly* é um tanto inconveniente como uma abreviação de *flyvemaskine* porque não pode ser pronunciada completamente como a primeira sílaba da palavra maior: a estrutura linguística do dinamarquês exige a adição de um golpe de glote à palavra monossilábica *fly* que a palavra maior não contém. O norueguês e o sueco, que não possuem essa dificuldade, adotam abreviações como essas com muito mais facilidade. Em tempos mais recentes, como é bem conhecido, temos observado a criação de um grande número de abreviações, principalmente em alemão, russo e inglês: *CDU* para *Christliche Demokratische Union*, *DDP* para *Deutsche Demokratische Partei*, *Komintérn* para *Kommunističeskij Inter-national*, *Komsomól* para *Kommunističeskij Sojúz Moloděži*, *UN* para *United Nations* (ONU “Organização das Nações Unidas”). Em português, entretanto, abreviações compostas pela sílaba inicial das palavras não são comuns (apesar de também podermos encontrar alguns exemplos desse tipo em organizações governamentais como ministérios, *Mcom* para *Ministério das Comunicações* e secretarias, *Secom* para *Secretaria de Comunicação Social*). A estrutura linguística do português prefere abreviações compostas pelas letras iniciais de cada palavra – as quais são comuns também em alemão e russo. Nesses casos, onde os exemplos podem produzir possíveis sílabas, algumas línguas têm uma tendência de pronunciá-las como tal: em francês pode-se pronunciar *URSS* como uma sílaba.

II. INTRODUÇÃO DE NOVOS SIGNOS

II. 1. Empréstimos

Um “empréstimo” (termo bem estabelecido, mas por vezes não tão claro) significa a transferência de um signo de uma língua para outra. Normalmente, palavras ou radicais de palavras são transferidos, mas outros tipos de signos também o podem ser, como prefixo *inter-*, o qual vem do latim, mas pode ser usado para construir palavras propriamente do português. *Palavras emprestadas* cuja forma externa produz um efeito de estranheza ou de incongruência – por exemplo, a organização silábica (*tsetse*) ou tipos especiais de acentuação – também são chamados de *estrangeirismos*. O que é característico dos estrangeirismos é que eles não seguem as regras da estrutura linguística da língua nativa, mas da sua língua de origem. Eles são como minorias que vivem sob suas próprias leis específicas. Consequentemente, eles estão sempre sujeitos a transformações causadas por falantes que desconhecem essa estrutura de origem e que

seguem, por outro lado, a familiaridade que tem com a língua nativa, ou a já citada etimologia popular (*rés-do-chão* para *rez-de-chaussée*). Já as palavras emprestadas que não se caracterizam como estrangeirismos, suas origens passam despercebidas aos olhos do homem comum e este ficaria surpreso ao descobrir que palavras como *batom* (do francês *bâton*), *álgebra* (do árabe *al-jabr*) ou *capim* (do tupi *ka'apii*) são palavras emprestadas.

Quando uma palavra é emprestada de uma fonte conhecida, é obviamente fácil reconhecer sua identidade; mas geralmente as funções-elemento não serão válidas. Já tivemos a oportunidade de referir-nos a palavras que foram levadas de uma língua à outra por esse meio: empréstimos gregos e latinos no alemão como *fraternize* (p. 34), *paternity* (p. 35), *pecuniary* (p. 35), *inspect* (p. 41), *pilgrim* (p. 65), *crocodile* (p. 68), *skeptic* (p. 68), dentre outras; empréstimos do francês no alemão como *barber* (p. 67) e *Muskedonner* (p. 71); empréstimos do italiano no francês como *choucroute* (p. 71). Podemos observar que enquanto as funções-elemento possuem a letra inicial em ale. *f* correspondendo ao lat. *p* (*father, fee*), os empréstimos do latim com *p*- inicial surgem em alemão como *p*- (*paternity, pecuniary*). Nesse caso, e na vasta maioria dos casos, há uma identidade entre a letra e o som na língua que empresta e na língua que toma emprestado. Ainda é preciso dizer que palavras emprestadas que deixam de ser estrangeirismos são adaptadas não apenas à estrutura linguística da língua de chegada, mas também ao uso linguístico desta. Essa adaptação é normalmente o primeiro passo rumo à assimilação de uma palavra emprestada na língua de chegada. Assim, por exemplo, pronunciamos *psicologia* como se houvesse um *i* após o *p* (uma adaptação à estrutura linguística do português) e pronunciamos a palavra *faux* (como em *faux pas*), emprestada do francês, com os sons usados na pronúncia de *foe* (adaptação ao uso linguístico do inglês)

Entretanto, existem instâncias – embora bem menos frequentes – quando uma palavra emprestada é transformada no momento do empréstimo de modo a estar em conformidade enquanto ou, geralmente, apenas em parte com as funções-elemento entre as línguas de empréstimo e de chegada (desde que elas seja, é claro, geneticamente relacionadas). Caso isso ocorra, a comunidade linguística deve conter falantes bilíngues o suficiente cuja familiaridade linguística os tornem consicentes das funções-elemento regulares entre as duas línguas. Por isso, existe uma consciência coletiva na dinamarca (sem dúvida influenciada pelos empréstimos do baixo alemão no dinamarquês) de que din. *sk* corresponde a alem. *sch* (pronunciado como [š]), como é ilustrado pelos seguintes pares *skabe* – *schaffen*, *skade* – *schaden*, e *sko* – *Schuh*. Consequentemente, a palavra

francesa *jus* “suco”, a qual é pronunciada como [žy] e foi levada ao alemão com a pronúncia [šy] por uma adaptação ao uso linguístico da língua alemã, surge em dinamarquês, o qual tomou-o emprestado do alemão, como *sky* “molho”.

Todavia, os falantes de uma língua, raramente são linguistas tão competentes ao ponto de levar a frente uma “fraude” como essa com sucesso. A familiaridade linguística normalmente estará preocupada somente com certas diferenças entre línguas que são particularmente marcantes para o ouvido. Na prática, é quase impossível detectar apenas pela forma externa de uma palavra se ela é emprestada ou não. De fato, a forma externa de uma palavra é o único critério de um linguista. Caso seja possível demonstrar que uma palavra esteja conectada a outra em uma língua diferente, mas falha em conformar-se às funções-elemento que foram estabelecidas na base de outros exemplos e não pode ser explicada de nenhuma outra forma (pela influência dos elementos, pela formação analógica ou pela introdução de um novo signo de alguma forma diferente, então deve-se concluir que tal palavra fora emprestada. O empréstimo então é uma explicação que sempre estará a mão quando outras explicações falharem. Entretanto, isso não significa dizer que criamos um depósito no qual colocaremos todos os casos irreduzíveis e os classificaremos como “empréstimos”. Como vimos, o empréstimo é algo bem preciso e bem definido, logo, esse processo é logicamente convincente quando outras explicações são impossíveis. Em todas essas instâncias em que podemos traçar detalhadamente as relações históricas e influências culturais que estão por trás do processo de empréstimo, podemos demonstrar que, sempre que encontramos uma palavra que não está geneticamente relacionada às palavras nas línguas geneticamente relacionadas e não explicável pela influência dos elementos, por formação analógica ou pela introdução de um novo signo de alguma forma diferente, então, de fato, a palavra *foi* introduzida a partir de uma instância exterior em um determinado período e por uma determinada rota. Existem, com certeza, muitos outros casos nos quais as relações históricas exteriores e as correntes culturais não são claramente reconhecíveis para permitir tal demonstração. Porém, nesses casos, devemos generalizar, como a ciência o faz em situações análogas: a partir do nosso conhecimento dos casos totalmente elucidados, podemos concluir que os casos menos conclusivos são do mesmo tipo, desde que nada contradiga essa afirmação. Na prática, então, a forma externa das palavras é nosso único critério decisivo para afirmar a existência de um empréstimo e, por esse meio puramente linguístico, somos capazes de fazer inferências sobre contatos que ocorreram na pré-história entre pessoas e correntes

culturais que, de outra forma, permaneceriam completamente ocultas porque não existem quaisquer outras fontes que poderiam suscitar essas informações.

Se por ventura deixemos de lado tais casos – especialmente aqueles mais recentes – em que as circunstâncias históricas externas são perfeitamente conhecidas de modo que a própria introdução da palavra por empréstimo possa ser datada e investigada, não existem quaisquer outros meios para estabelecer um empréstimo senão o interno, puramente linguístico. Em princípio, não existem limites para o que pode ser tomado de empréstimo de uma língua para outra. Normalmente um signo será naturalmente tomado de empréstimo quando acompanha a coisa designada: um artigo de exportação carrega consigo sua designação, como quando o lat. *uīnum* “vinho” se espalhou como uma palavra de empréstimo por quase toda a Europa. Mas esse é apenas um caso muito específico. Uma palavra não é meramente tomada de empréstimo porque designa alguma coisa que não possui um desígnio nativo. Os falantes tomam emprestadas palavras por um desejo de imitação e tais empréstimos estão sujeitos a toda sorte de caprichos de estilo. As pessoas farão isso porque a palavra estrangeira soa mais fina ou mais bela ou ainda mais divertida ou mais interessante e ninguém pode prever o que as pessoas considerarão mais fina, bela, divertida ou interessante. Populações bilingues podem também, é claro, introduzir empréstimos em uma língua simplesmente porque possuem uma certa dificuldade em manter os sistemas das duas línguas conceitualmente separados. Exemplos disso podem ser encontrados em todas as regiões fronteiriças. Porém, novamente, trata-se de um caso muito específico. Antes do processo de empréstimo, é impossível prever o que poderá e o que não poderá ser emprestado. As palavras que podem passar por esse processo em uma língua podem chegar a um número formidável. Verificou-se que no albanês pode-se localizar apenas seiscentas palavras que não sejam fruto de empréstimo, com todo o resto vindo do latim, línguas românicas, línguas eslavas, grego e turco. Até mesmo o grego, que é considerada uma das línguas indo-europeias mais puras e que exerce um papel maior que qualquer outra língua nos estudos comparativos do indo-europeu, contém apenas um pequeno número de palavras que podem ser geneticamente consideradas da base do indo-europeu. Pode-se presumir, portanto, que as palavras gregas são empréstimos de outras línguas, majoritariamente, quiçá, de línguas não indo-europeias. Assim como vimos acima que algumas comunidades linguísticas estão mais inclinadas que outras a admitir abreviações de signos, podemos dizer a mesma coisa em relação aos empréstimos, sem que possamos falar qualquer coisa sobre as causas a que se deve esta disposição particular deste ou

daquele grupo linguístico. Existem línguas nas quais as muitas palavras internacionais de origem greco-latina que abundam na maioria das línguas europeias modernas têm muita dificuldade em serem admitidas, porque a comunidade linguística prefere utilizar o léxico nativo para designar o que estiver em pauta. Assim, *universidade* em islandês é *háskóli* “escola de ensino médio”, em finlandês, *yliopisto* “instituto de altos estudos” e, em húngaro, *egyetem*, um derivado artificial do numeral *egy* (pronunciado como [ed’]) “um”, em alusão ao latim *ūniversitās*, um derivado do numeral *ūnus*; enquanto que outras línguas europeias utilizam a palavra que aparece em português como *universidade*. O islandês possui uma fama em particular (ou, se preferir, notória) por sua teimosia em adotar empréstimos linguísticos.

Dessa maneira, não podemos prever o que poderá ser tomado de empréstimo, mas podemos afirmar a possibilidade de que tudo pode passar por esse processo. De um lado, não podemos estabelecer de antemão se uma língua resistirá ao processo de empréstimo ou, de outro, se aceitará. Não podemos também estabelecer leis gerais que indiquem a direção em que esses empréstimos tomarão. Poderão passar de um status social preterido a um que o seja preferido ou vice-versa, de línguas com poucos falantes a línguas muito populares ou vice-versa, da língua de povos subjugados à língua de seus governantes ou vice-versa, de um dialeto à língua padrão ou da língua padrão a um dialeto. Onde quer que haja um encontro entre duas comunidades, haverá a possibilidade de empréstimo linguístico.

Ao estabelecermos o empréstimo linguístico como um fato, então, o linguísta vê-se com apenas o critério linguístico como ferramenta, do qual o mais importante é o fato de que as palavras envolvidas constituem contraexemplos às funções-elemento. Vimos acima que em celta a uma correspondência entre um morfema zero e o I.E. **p* e o lat. *p*, e que, por exemplo, o irlandês possui uma correspondência com o ing. *fish*, lat. *piscis* a partir da forma *iasc* (p. 40). Quando verificamos que a palavra que corresponde a *fish* é *pysg*, concluímos então que trata-se de um empréstimo do latim. Nesse caso, como percebe-se, essa afirmação pode encontrar suporte em muitos exemplos, porque sabemos que em geral existiu uma influência considerável do latim no País de Gales. Comumente pode ocorrer uma predominância de empréstimos em uma determinada direção entre uma língua e outra. Um movimento geral de palavras emprestadas que exerceram um importante papel na Europa vem dos países do mediterrâneo, especialmente aqueles de língua semítica, e passa pelo grego e então pelo latim, do latim ao francês e, finalmente, ao germânico. Em inglês, a palavra *sack* chegou até nós através de um longo caminho. O

dinamarquês, como vimos, adquiriu um grande número de empréstimos, alguns do francês e outros do escandinavo. Durante um longo período de tempo, o francês tomou emprestado um vasto léxico do latim. Vilhelm Thomsen demonstrou (parcialmente em sua tese de 1869) que o finlandês – a qual vimos ser uma língua bastante conservadora em tempos passado em relação a empréstimos de outras línguas – deve, em tempos pré-históricos, ter aceitado um número considerável de palavras emprestadas, especialmente de línguas germânicas. Alguns desses empréstimos são mais próximos em sua forma à estrutura do germânico comum do que qualquer outra coisa que fora registrada nas línguas germânicas. Assim, esses empréstimos possuem relevada importância na linguística comparativa germânica.

II. 2. Tabu

A palavra *tabu* é polinésia e refere-se a uma situação em que uma palavra ou nome só pode ser usada em uma comunidade sob condições especiais, seja porque somente algumas pessoas podem proferi-la, seja por alguma restrição circunstancial. Esse fenômeno foi primeiramente observado entre comunidades tribais, cujas noções religiosas tornam esse fenômeno bastante difundido. Entretanto, em investigações mais recentes, têm-se demonstrado que o tabu é reconhecível em todas as sociedades em qualquer período de tempo e desempenha um papel considerável em nossas próprias línguas. O exemplo mais óbvio disso é o fato de que é considerado impróprio nomear diretamente fenômenos conectados com a vida sexual ou com os processos digestivos. Esse tabu, como pode vir-se a pensar, não é um fenômeno global. Está ligado, na verdade, a certas sociedades e é desconhecido em muitas fora do contexto moderno europeu; não existia na Grécia antiga, por exemplo. Porém, outras áreas da vida – a área religiosa, por exemplo – também pode estar cercada por tal mistério linguístico: não se pode proferir o nome de Deus em vão. Consequentemente, se alguém precisa mencionar tais coisas, ele deve ou usar alusões ou circunloquções ou ainda simplesmente mudar as palavras para designá-las. Muitas vezes, como é bem conhecido, uma palavra estrangeira é usada: não é a coisa em si mesma que está sob o tabu, mas o signo, e se alguém utiliza um signo estrangeiro, o estorvo desaparece. Ou ainda: outra palavra é arbitrariamente escolhida, mesmo significando outra coisa bem diferente, mas cuja forma externa aproxima-se daquela que de maneira suficiente que possa lembrá-la (*sewer-pipe* em vez de *suicide*, em inglês). Ou, finalmente – e isso será de grande interesse para nós – a palavra é arbitrariamente

transformada. Isso pode ser feito de diferentes formas: por abreviação (por exemplo, TB pode ser utilizada em lugar de *tuberculose* para que não se evoque a temida doença a partir de seu nome), ou por transmissão de elementos (em dinamarquês *pokker* “o diabo”, se derivado de *kopper* “varíola”, só pode ser explicado a partir do entendemos por tabu e não como uma metátese, uma vez que o processo observado vai contra a lei geral que rege a metátese; veja a p. 68), ou por substituições de diferentes elementos em certos lugares na palavra. Pode-se talvez, a título de exemplo, observar que pode ser considerado admissível em dinamarquês chamar um pequeno negócio pelo verbo *nisse*. Essa palavra usa um *n* para substituir o *t* de outra palavra, a qual está sob forte influência de um tabu, apesar de que a palavra com *t*, por sua vez, fora originalmente introduzida como uma substituta arbitrária de uma terceira palavra, a qual (como a palavra correspondente em português) está sob um tabu ainda mais forte. Dessa forma, o tabu produz contraexemplos às funções-elemento, Pode existir ainda um reforço mútuo causado pelo tabu e pela influência dos elementos. A transformação de lat. *meretrīx* “prostituta” em *meletrīx* (**meltrīce* é pressuposto por *meautris*, do francês antigo e por *meltriz*, do provençal) pode ser explicada por uma dissimilação como aquela observada em *pelegrīnus*; mas a forma *menetrīx*, que também pode ser observada, não pode ser explicada pela estrutura sonora do latim e, portanto, configura-se como um tabu.

II. 3. Neologismo

Por fim, devemos observar que, obviamente, também é possível introduzir signos completamente novos em uma língua, mesmo que eles não sejam nem transformações nem derivações baseadas em outros signos, mas que ainda assim sigam as regras de formação silábica da estrutura linguística. Novamente, diferentes comunidades linguísticas, em diferentes períodos, comportaram-se de modos distintos em relação a esse processo. A maioria das línguas da Europa ocidental, quando comparadas a outras regiões linguísticas, possuem uma tendência muito baixa em aceitar neologismos. Há em nossas línguas, na verdade, uma aversão aos neologismos, com exceção às onomatopeias e à licença poética que é dada aos escritores. Por outro lado, nem sempre esse é o caso. Um exemplo marcante disso em tempos mais recentes foi a introdução arbitrária de milhares de novas palavras no estoniano por um decreto governamental no século vinte, palavras estas que foram criadas, por assim dizer, do nada, reunindo elementos de expressão estonianos de novas maneiras, de acordo com as regras de

formação de sílabas dessa língua. Tais neologismos, é claro, nunca poderão ser mais que contraexemplos aparentes às funções-elemento. Uma palavra recém-criada pode deixar perplexo o investigador da relação genética apenas se for bastante próxima em forma do que ele pode ser levado, pelas funções-elemento, a esperar encontrar na língua. O linguista deve, portanto, também estar atento a essa possibilidade. No mais, o linguista deve ter em mente que existem relativamente poucas palavras em uma língua (como em grego, por exemplo) que podem ser explicadas geneticamente. Isso nem sempre será causado pelos empréstimos linguísticos, mas também pela criação de neologismos.

Famílias linguísticas

O método que delineamos tornou possível o reconhecimento de um grande número de famílias linguísticas ao redor do mundo. Entretanto, como vimos, existem famílias linguísticas que comportam diferentes graus de parentesco, então, para propriamente designá-las, usaremos os seguintes termos: reservaremos o termo *família* para designar um grande grupo de línguas como as que compõem a família do indo-europeu. Tal família será então dividida em *grupos*, *subgrupos*, *ramos* e assim por diante. Procederemos agora a uma revisão sumária das mais importantes famílias linguísticas estabelecidas até o momento.

I. A família indo-europeia (veja a p. 33)

1. O grupo *germânico*

A. O subgrupo *germânico oriental*

- a) O ramo *gótico*, a qual inclui diversas línguas, mas apenas uma possui registros bem preservados (veja a p. 33).
- b) O ramo *escandinavo* (ou *nórdico*). O *nórdico primitivo* é o mais antigo estágio dessa ramificação. Em suma, é a língua mais antiga e arcaica de todas as línguas germânicas como é observado nas runas que foram encontradas. O estágio mais recente divide-se em duas sub-ramificações: o *escandinavo ocidental* (*islandês*, *feroês*, *norueguês*) e *escandinavo oriental* (*sueco*, *dinamarquês*). De 1530 a 1814, o dinamarquês foi a língua padrão e a língua literária da Noruega. Em norueguês, “*riksmål*” ou “*bokmål*” são palavras dinamarquesas com um toque norueguês; ao *landsmål*, uma língua padrão formada na base do dialeto escandinavo ocidental de *Søndmøre*, foi dado igualmente o status de oficialidade em conjunto com o “*riksmål*” desde 1892 e provê o as bases para o “*nynorsk*” (novo norueguês). Em 1938, uma terceira língua padrão, o “*samnorsk*” (norueguês comum), fora introduzida como um meio-termo.

B. O subgrupo *germânico ocidental*

- a) O ramo *anglo-frísio* (*inglês*, introduzido na Inglaterra por invasores vindo do continente e difundido em todo o Império

Britânico e nos Estados Unidos; *frísio*, falado ao longo da costa ocidental da Holanda do Norte e da Alemanha).

- b) O ramo holandês-alemão (*holandês*, com duas línguas escritas: o holandês e o neerlandês, na Holanda, na Bélgica e nas colônias e o africâner, na África do Sul; *baixo-alemão* e *alto-alemão* – veja a p. 46)

Dentro do indo-europeu, as línguas do grupo germânico englobam o maior número de falantes – por volta de 1,5 bilhão²⁴, dos quais a maioria são falantes de inglês, a língua mais difundida no mundo.

2. O grupo *celta*

A. O subgrupo *gaélico* (*irlandês*; *escocês gaélico*, introduzido na Escócia por imigrantes irlandeses no século quinze; *Manx*, hoje extinto, fora falado na Ilha de Man)

B. O subgrupo *britânico*

a) O ramo *continental* (*gaulês*, confirmado por inscrições e extinto depois da conquista da Gália e a criação do Império Romano).

b) O ramo *insular* (*galês*, no País de Gales; *córnico*, falado em Cornwall e extinto; *bretão*, introduzido na Britânia pelos imigrantes vindos da Grã-Bretanha nos séculos V e VI; o bretão possui quatro línguas escritas diferentes).

3. O grupo *itálico*

A. O subgrupo *itálico ocidental*, consideravelmente mais difundido que o oriental no início da história.

a) O ramo *umbro-samnita* (*umbro*, falado na região italiana da Úmbria; *samnita*, fora falado na região sul da Itália, com a língua oficial chamada de *osco*).

b) O ramo *lepôntico* (na região noroeste dos Alpes).

B. O subgrupo *itálico oriental*

a) O ramo *siciliano*

b) O ramo *continental*, o qual inclui, dentre outras muitas línguas de baixa difusão, o *latim*, no extremo ocidente do Lácio, a partir da

²⁴ N.T.: Número atualizado com base em...

nascente do rio Tibre e desembocando na cidade de Roma. Com exceção do latim, que floresceu assim como outras línguas o fizeram, todas as línguas itálicas morreram antes do fim da antiguidade e são conhecidas apenas por manuscritos. No início dos tempos históricos, o latim fora falado em uma área menor que a ilha de Zelândia na Dinamarca e, no fim das contas, veio a dominar metade de um continente. No fim da antiguidade, foi continuada pelas línguas *românicas*, as quais podem ser separadas em seis sub-ramificações: (1) *íto-sardenho* (*italiano*, falado na península itálica, na Sicília e, em certo grau, em Córsega; *sardenho*, falado na Sardenha e, até certo ponto, também em Córsega); (2) *hispano-romance* (*espanhol*, falado na maior parte da Espanha, noroeste da África, em toda a América Latina – exceto no Brasil – e nas ilhas Filipinas; *português*, falado em Portugal, Brasil, Angola, Moçambique, Guiné-Bissau, Guiné Equatorial, Cabo Verde, São Tomé e Príncipe, Timor-Leste e na região autônoma de Macau na costa sul da China; *galego*, no noroeste da Espanha); (3) *galo-românico* (*Catalão*, falado principalmente na Catalunha e no sudeste da França; *provençal*, do sul da França; *francês*, a língua oficial da França, Benin, Burkina Faso, Camarões, dentre outras antigas colônias, além de ser a língua oficial da Bélgica (em conjunto com o neerlandês), da Suíça (em conjunto com o alemão, o italiano e o reto-romanche), do Canada (em conjunto com o inglês) e de Luxemburgo (em conjunto com o alemão); (4) *reto-romanche* (no sudeste da Suíça e no nordeste da Itália); (5) *dalmático* (falado na Dalmácia mas extinto desde o século XIX) e (6) *romeno* (falado na Romênia, com exceção do extremo sudeste da Transilvânia e difundido na península balcânica).

4. O grupo *helênico*

- A. O subgrupo *macedônio*, compreendendo apenas uma língua extinta e mal documentada.
- B. O subgrupo *grego*, também compreendendo apenas uma língua.

5. O grupo *báltico*

- A. O subgrupo *báltico setentrional* (*lituano* e *letônio*). O lituano é a única língua viva que é explicável diretamente a partir do sistema de fórmulas comuns ao indo-europeu.
 - B. O subgrupo *báltico meridional*, que compreende diversas línguas, das quais apenas uma sobreviveu: o *prussiano* que morreu ainda no século XVII, desbancado pelos falantes de alemão que se apossaram do nome de prussianos bálticos.
6. O grupo *eslavo*
- A. O subgrupo *eslavo meridional*
 - a) O ramo *búlgaro* (apenas uma língua) e *macedônio* (também apenas uma língua, da qual conhecemos uma antiga forma chamada *antigo eslavo eclesiástico* e é a mais antiga língua eslava registrada).
 - b) O ramo *iugoslavo* (*servo-croata* no Leste e *esloveno* no Oeste).
 - B. O subgrupo *eslavo setentrional* (*sérvio* – também conhecido como *lusaciano* ou *wendish* – com duas formas literárias: a língua dos wends na baixa e na alta Lusácia; *polábio*, falado nos arredores do rio Elba (*Laba* em eslavo) e em Hanover, está extinto desde o século XVIII; *polonês*; *tcheco* e *eslovaco*).
 - C. O subgrupo *russo* – ou *eslavo oriental* – (*grande russo*, *ucraniano* e *bielorrusso*).
7. O grupo *albanês* (constituído por apenas uma língua).
8. O grupo *armeno* (constituído por apenas uma língua).
9. O grupo *iraniano*
- A. O subgrupo *iraniano ocidental* (*persa* em três estágios: *persa antigo* registrado em inscrições cuneiformes, *pálavi*, e o *persa moderno*, a língua oficial do Irã; o *avéstico*, que era nativo, em tempos antigos, à região norte da Pérsia antiga, fato que é atestado pelas inscrições de teor religioso deixadas pelo líder Zaratustra. Além disso, está intimamente relacionado ao mais moderno *pálavi caldeu* que, mais tarde, fora extinto; *curdo*, nos arredores da região montanhosa na fronteira entre a Turquia e o Irã; dentre outras muitas línguas, incluindo algumas nas várias regiões do Irã).
 - B. O subgrupo *iraniano oriental* (*sogdiano*, atestado pelos textos vindos do Turquestão oriental (ou chinês), teve continuidade no yaghnobi, que é falado no nordeste de Pamir, a Leste do Mar de Aral; *sakico*, registrada

em textos encontrados também Turquestão oriental; *afegão*, ou *pastó*; *osseto*, falado no Cáucaso; e outras várias línguas, incluindo aquelas faladas na região de Pamir e do Baluchistão).

10. O grupo *hindi*. O *hindi antigo* é formado por dois dialetos: o *védico* (encontrado nos Vedas) e o *Sânscrito*. As línguas *médio-indianas* são conhecidas sob o nome geral de *prácrito* e inclui o *páli*, língua sagrada do Budismo meridional. O *indiano moderno* é constituído por um grande número de línguas (incluindo o *romani*, a língua dos Ciganos), que são faladas por 700 milhões de pessoas. Atualmente, o hindi fica atrás apenas do chinês, do inglês e das línguas românicas com relação ao número de falantes; em seguida, temos o árabe com cerca de 274 milhões.
11. O *hitita*, uma língua antiga descoberta apenas recentemente e registrada parcialmente na forma de hieróglifos em escrita cuneiforme; foi nativa à região da Ásia menor (ou Anatólia) no século quinze antes de Cristo.
12. O *tocariano*, descoberto no início do século vinte e falado no Turquestão oriental no século sete depois de Cristo.

À família indo-europeia, pertencem ainda alguns remanescentes linguísticos (ou seja, línguas mortas que foram transmitidas de forma fragmentária) da região mediterrânea: *messápico*, *venético*, *trácio*, *frígio* e *lício*.

II. A família *hamito-semítica*

1. O grupo *hamítico*

- A. O subgrupo *egípcio*, registrado em hieróglifos de até 4 mil anos a.C. Seus estágios iniciais são o *egípcio demótico* e o *cóptico*, os quais morreram e foram substituídos pelo árabe no século dezesseis d.C.
- B. O subgrupo *líbico-berbere* (extinto na Líbia e substituído, majoritariamente, pelo árabe; ainda é falado na Argélia, Tunísia, Marrocos e na região do Saara).
- C. O subgrupo *cuxítico* (no oriente da Núbia, Etiópia, Somália e Leste africano).

2. O grupo *semítico*

- A. O subgrupo *acadiano*, consistindo de apenas uma língua – fora falada em tempos antigos na região do Iraque e registrada em escrita cuneiforme.
- B. O subgrupo *semítico ocidental*

- a) ²⁵ O ramo *semítico do Noroeste* é formado por duas subramificações: *cananeu* (*cananeu antigo, moabe, hebraico, fenício*) e *aramaico*. Com exceção do hebreu e alguns dialetos do aramaico oriental, todas essas línguas estão extintas. O hebraico, com certeza, fora extinto na Palestina como uma língua falada no terceiro século a.C. e foi substituído pelo aramaico, mas continuou a ser uma língua científica e religiosa e que foi continuada, de certa forma, pelo iídiche, que pode ser visto como uma junção do hebraico e do alto alemão. Além disso, o hebraico constitui um caso especial de retorno, pois foi reintroduzido como uma língua viva em Israel e é falado por cerca de 8 milhões de pessoas no mundo atualmente.

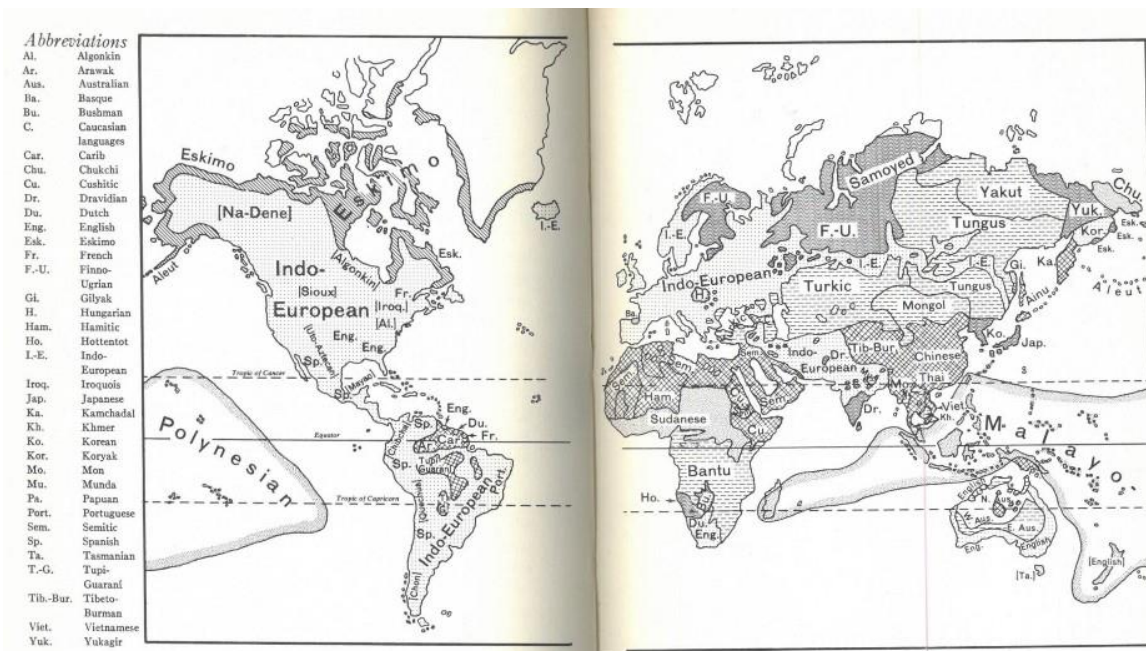


Figura 1 - Reprodução do mapa apresentado nas páginas 74 e 75.

²⁵ N.T.: Tradução das abreviações listadas no mapa: Abreviações: Al. Algonquim; Ar. Aruaque; Aus. Australiano; Ba. Basco; Bo. Bosquímano; Car. Caribe; Chu. Chukchi; Co. Coreano; Cor. Coriaco; Cux. Cuxítico; Dr. Drávidico; Esp. Espanhol; Esq. Esquimó; F.-u. Fino-úgrico; Fr. Francês; Gi. Gilyak; H. Húngaro; Ham. Hamítico; Ho. Hotentote; Hol. Holandês; I. -e. Indo-Europeu; Ing. Inglês; Iroq. Iroquês; Jap. Japonês; Ka. Kachamdai; L. c. Línguas caucasianas; Mo. Mom; Mu. Munda; Pa. Papua; Port. Português; Que. Quemer; Sem. Semítico; T. -g. Tupi-guarani; Ta. Tasmânio; Tib. -bir. Tibeto-birmanês; Viet. Vietnamita; Yuk. Yukagir.

b) O ramo *semítico do Sudoeste* é formado por três sub-ramificações: *árabe* (que, com o advento da cultura islâmica, é a língua da família hamito-semítica mais difundida; e, ao lado do inglês e do chinês, é uma das línguas culturais modernas mais difundidas; é falada atualmente na Arábia Saudita, Iraque, Síria, Egito, no noroeste da África e em Malta), *árabe meridional* e *etíope* (que inclui sete línguas modernas, das quais a mais conhecida é o *amárico*, língua oficial da Etiópia).

III. A família *banta* (que vem de *ba-ntu*, plural de *mu-ntu* “ser humano”), que é formada por um grande número de línguas nativas de toda a parte sul da África com exceção do sudoeste na região da Namíbia.

IV. A família *urálica*

1. O grupo *samoieda*, formado por cinco línguas que se estendem pelo lado setentrional do que já fora a União Soviética. Duas delas adquiriram o status de línguas literárias em 1930.

2. O grupo *fino-úgrico*

A. O subgrupo *úgrico*

a) O ramo *ob-úgrico* (*ostyak*, nos arredores do rio Ob; *mansi* ou *vogul* ao oeste, próximo aos Montes Urais; ambas as línguas adquiriram status literário em 1930 .

b) O ramo *hungáro* (apenas uma língua, que é a oficial da Hungria; também é falada em algumas esparsas regiões da Eslováquia e da antiga Iugoslávia.

B. O subgrupo *lapônico* (contém apenas uma língua; mas inclui seis distintos grupos de dialetos na Noruega, Suécia, Finlândia e na região da antiga União Soviética.

C. O subgrupo *fino-permiano*

a) O ramo *permiano* (*komi* ou *zyriano*, na região da República de Komi, localizada no nordeste da Rússia europeia, com uma rica literatura produzida no século quatorze; *udmurte* ou *votyake*, na República da Udmúrtia, ao nordeste de Cazã.

- b) O ramo *fino-mordoviano*, formado por duas sub-ramificações: (1) *mordoviano*, uma língua falada em esparsas regiões ao redor do rio Volga entre as cidades de Nijni Novgorod e Kuybyshev; (2) *mari* ou *cheremisse*, formada por apenas uma língua falada na República de Mari El (ao norte de Cazã) e em áreas esparsas República da Udmúrtia; *finlandês báltico*, formado por dez línguas muito próximas, das quais conhece-se bem o *finlandês*, *carélio* e o *estônio*.

V. A família *altaica*

1. O grupo *turco*, formado pelo *turco*, ou *otomano*, na Turquia, no norte da Síria e em algumas regiões ao norte da Grécia e da Sérvia; inclui ainda algumas outras línguas que perpassam a região da antiga União Soviética – incluindo-se aqui a língua oficial do Azerbaijão, da República do Tartaristão, da República da Chuvachia e da República de Sakha ao longo da foz do rio Lena – como também da província de Sin-Kiang no Turquestão oriental. O *turco antigo*, que fora recuperado a partir das inscrições de Orcom descobertas na Mongólia, foi decifrado por Vilhelm Thomsen em 1893 e tornou-se a base para a comparação genética das línguas turcas.
2. O grupo *mongol*, formado por línguas majoritariamente presentes na Mongólia e na região adjacente à Sibéria em conjunto com partes da província de Kuku-nor. A literatura produzida nessas línguas data do século treze.
3. O grupo *tungúsico* (*evenki*, falado na maior parte da Sibéria e ao Leste do rio Lenissei. É considerada uma língua literária desde 1930. *Manchu*, falado em Manchukuo e em Dal'neVostochny Kray (entre Nikolayevsk e Vladivostok) com sua literatura iniciando no século doze).

I. A família *sino-austronésica*:

1. O grupo *thai-chinês*, que inclui o *chinês*, língua que é falada pelo segundo maior grupo de pessoas no mundo (1,1 bilhão), e as línguas *tailandesas* da Indo-China, das quais a mais conhecida chama-se *siamês* ou *tailandês* e é a língua oficial da Tailândia.
2. O grupo *tibeto-birmanês*.
3. O grupo *austronésico*:

- A. O subgrupo *malaio-polinésio* (encontrado em algumas ilhas do Pacífico, Malásia, Madagascar, Indonésia; em tempos passados, foi a língua cultural de Java).
- B. O subgrupo *austro-asiático*, que é formado por três ramos: as línguas *munda* da Índia, encontradas especialmente na província de Bihar (a mais conhecida dentre essas línguas é o *santáli*), as línguas *mon-khmer* da Indochina e o *vietnamita*.

Embora tenhamos usado *família* como o termo mais inclusivo para uma ampla classe de línguas geneticamente relacionadas, não há nada, é claro, que impeça que uma determinada família seja relacionada a outras famílias linguísticas. Deve-se observar que é possível provar que duas línguas são geneticamente relacionadas, mas é impossível provar que *não* o são. Portanto, usamos o termo *família* para designar uma ampla conjunto de línguas que, a partir dos dados obtidos a certa altura da pesquisa, podem ser agrupadas em uma mesma classe. Nesse aspecto, família torna-se um termo relativo. Caso classifiquemos duas famílias linguísticas como grupos separados e, mais a frente na pesquisa, verificarmos que essas famílias são geneticamente relacionadas, devemos, então, classificá-las como uma única família, cada uma pertencente a um grupo diferente e assim por diante. Porém, normalmente esse processo não seria prático. Embora, por exemplo, saibamos agora que as línguas indo-europeias são geneticamente relacionadas a outras línguas, ainda continuamos a nos referir a esse grupo como a família indo-europeia, e estamos certos de assim fazê-lo, desde que admitamos a possibilidade de existirem *famílias de famílias*.

Em cada caso, será meramente uma questão prática quer mudemos nossos termos ou não. Por um longo tempo, o que chamamos de família sino-austronésica era classificada como cinco famílias diferentes que, até aquele ponto, não demonstravam qualquer relação genética: (1) sino-tibetano, que inclui o thai-chinês e o tibeto-birmanês, (2) malaio-polinésio, (3) munda, (4) mon-khmer, (5) vietnamita. Em 1889, o linguísta alemão Ernst Kuhn demonstrou que as últimas três eram relacionadas e em 1908-10, o linguísta dinamarquês Kurt Wulff foi capaz de demonstrar a relação dessas ao malaio-polinésio. Muito tempo depois, a relação genética entre o austronésico e o sino-tibetano foi demonstrado em um trabalho póstumo de Wulff. Construimos nosso processo a partir desses resultados e os combinamos em uma única família, porque as relações genéticas apresentadas por Wulff contradizem a antiga classificação: a relação entre o thai-chinês e

o austronésico é mais estreita do que a relação entre aquele e o tibeto-birmanês, então não há nenhuma base que sustente a manutenção de uma outra família sino-tibetana.

Uma relação genética entre o indo-europeu e o hamito-semítico foi demonstrada em detalhe pelo linguísta dinamarquês Hermann Møller através do método das funções-elemento em estudos entre os anos de 1906 e 1917. Mais tarde, pelo mesmo método, o africanista francês Mlle Liliás Homburger pensou ser possível demonstrar uma relação genética entre o hamito-semítico e o banta e, ainda, conjecturou adicionar a ambas as famílias algumas línguas do nordeste (nilótico) e do oeste do Sudão, das quais a primeira parece ser mais próxima do cóptico do que o egípcio antigo e o demótico. Também não há dúvida da relação entre o indo-europeu e o urálico depois que Holger Pedersen conseguiu demonstrar com sucesso algumas correspondências entre elementos.

Ao combinar esses resultados de amplo espectro, poderemos talvez reduzir as seis famílias acima a somente três: (1) *nostrático* (nome proposto por Holger Pedersen para aquelas línguas relacionadas às nossas próprias, do lat. *nostrās* “nossos queridos conterrâneos”), incluindo, até onde podemos vislumbrar atualmente, o indo-europeu, o hamito-semítico, o urálico e, possivelmente, o banta (em conjunto com as línguas sudanesas citadas acima); (2) *sino-austronésico*; (3) *altaico*.

As famílias linguísticas abordadas aqui são as únicas classes de línguas que possuem um tamanho considerável para ser estabelecidas como famílias pelo método das funções-elemento. O mesmo método fora aplicado às famílias de famílias das quais falamos anteriormente, mas como essa grande síntese ainda faz parte de um estudo recente, cabe a nós, prudentemente, admitir que as provas para isso ainda estão no porvir, mesmo que saibamos que isso é apenas uma questão de tempo.

Tomadas em conjunto, as línguas que listamos acima são faladas por aproximadamente 90% da população mundial atual. Mas a linguística deve, é claro, preocupar-se também com as línguas mortas assim como com as vivas.

As línguas restantes do mundo constituem um enorme conjunto de comunidades linguísticas geneticamente isoladas ou famílias que possuem um tamanho restrito. Exemplos disso são o japonês, o coreano, o ainu – no norte do Japão²⁶ –, as línguas paleoasiáticas do nordeste da Sibéria (uma classificação puramente geográfica), o caucasiano meridional (uma família de línguas), o caucasiano setentrional (uma classificação mais geográfica que linguística para um grande número de línguas espalhadas por essa região),

²⁶ De acordo com pesquisas mais recentes (Greenberg, 2000), o ainu seria uma língua indo-europeia.

dravídico (uma família de línguas do sul da Índia), as línguas australianas (uma mera classificação geográfica), as línguas não-banta do Sudão, a família khoisan no sudeste da África (bushman e hotentote), o basco (uma língua isolada na fronteira entre a França e a Espanha), o esquimó (em conjunto com o aleúte) e, por fim, o extraordinário e extenso grupo das línguas ameríndias, que são classificadas em não menos que algumas centenas de diferentes famílias. A esse de línguas geneticamente isoladas (ou de famílias), poder-se-á futuramente adicionar alguns remanescentes linguísticos, principalmente das regiões do Oriente Próximo e dos países mediterrâneos (por exemplo, o sumério no Iraque e o Etrusco na Itália).

Línguas originárias

Algumas páginas acima, demos uma definição operacional da relação genética (p. 50) e demonstramos que ela consiste do que chamamos de elementos-funções, ou seja, correspondências constantes entre os sistemas dos elementos da expressão nas línguas relacionadas. Uma relação genética entre línguas é evidenciada pela demonstração de tais funções-elemento e de nenhuma outra forma. A similaridade superficial entre línguas ou entre signos em diferentes línguas não tem nenhum papel seja na relação genética, seja em sua demonstração. Tal “similaridade” não pode ser utilizada como um critério científico, porque uma sensação de que determinadas coisas são similares é puramente uma questão subjetiva (veja acima na p. 62). Em muitas instâncias, pode-se demonstrar que dois signos são geneticamente relacionados mesmo que a maioria das pessoas não veja nenhuma semelhança entre eles. Essas formas, tais como o nominativo singular *hayr* em armênio e *patér* em grego, ou o genitivo singular *hawr* em armênio e *patrós* em grego, provavelmente não seriam vistas como semelhantes a partir de sua aparência externa, embora possam, a partir das funções-elemento, serem vistas como idênticas. Por outro lado, dois signos podem ter uma forte semelhança sem possuir nenhuma relação genética de qualquer tipo. Exemplos primários disso são a maioria das palavras emprestadas e, se tentarmos basear nossa perspectiva na semelhança externa, seria impossível distinguir entre palavras emprestadas e aquelas que são geneticamente relacionadas. Mas, é claro, também existem casos de dois signos em línguas diferentes que se parecem por puro acaso: “grande” é *dìdis* em lituano e *didi* em georgiano, mas as duas línguas não possuem qualquer relação genética ou de qualquer outro tipo; “to call” é *kaleîn* em grego, mas essas não podem ser a mesma palavra, já que em grego *k* nessa posição corresponde a ger. *h* (cf. *kúōn* e *hound*, *he-katón* e *hund-red* etc).

Agora, se questionarmos o porquê da existência dessas funções-elemento entre duas línguas, que chamamos aqui de relações genéticas, concluiremos que isso só pode ser verdadeiro se concordarmos que ambas as línguas possuem uma origem em comum. Essa é a única explicação possível, então, somos logicamente compelidos a aceitar essa consequência das relações genéticas. Tais relações consistem em uma conexão entre os elementos da expressão e os signos de duas línguas e a razão para essa conexão deve ser necessariamente que os elementos da expressão e os signos dessas duas línguas são desenvolvimentos posteriores de formas antigas que eram compartilhadas pelas duas línguas, ou, em outras palavras, que essas duas línguas se desenvolveram a partir da

ramificação de uma *língua base* (mais conhecida como uma *língua originária*). Desse modo (de acordo com a classificação dada no capítulo anterior), precedendo o dinamarquês e o sueco, deve haver um escandinavo oriental originário, precedendo o escandinavo oriental e o escandinavo ocidental deve haver também uma língua escandinava originária, antes dessa língua e do gótico, deve haver uma língua germânica oriental originária, antes dessa língua e do germânico oriental, uma língua germânica originária, antes dessa e das outras línguas dos grupos indo-europeus, uma língua indo-europeia originária e antes dessas e das famílias linguísticas geneticamente relacionadas ao indo-europeu, uma língua nostrática originária. Agora temos uma base sólida e real embaixo de nossos pés e, assim, podemos suplementar nossa definição operacional anterior de relação genética com uma *definição “real”*: *relações genéticas entre duas línguas significa que ambas se desenvolveram a partir de uma mesma língua base/originária*.

Tudo o que sabemos com certeza sobre a língua originária é o que podemos ler a partir das fórmulas que estabelecemos a partir da comparação genética entre línguas. De forma alguma a língua originária nos fora transmitida, nós a reconstruímos a partir das funções-elemento. A língua originária que pressupomos ser ascendente às línguas românicas pode ser chamada de latim, porque podemos concluir a partir de nosso conhecimento histórico que *deve* ter sido o latim que, por ramificações, desenvolveu-se nas línguas românicas. Porém, como já havíamos notado (p. 44), a língua originária que reconstruímos para as línguas românicas não é o mesmo latim que conhecemos a partir da literatura transmitida até nós.

Os elementos da expressão e partes de palavras (quase nunca palavras inteiras; veja na p. 48) que estabelecemos em nossas fórmulas devem, então, ser atribuído à língua originária. Eles são fragmentos de uma língua que não poderíamos conhecer de outra maneira. E aqui será de grande importância relatar, da forma mais clara possível, sobre o conteúdo real que estamos autorizados a incluir em nossas fórmulas ou, em outras palavras, o que temos certeza e o que não temos certeza sobre as línguas originárias.

A. O que temos certeza:

1. Que tais línguas existiram;
2. Que essa língua contém em sua estrutura os elementos da expressão representados por nossas fórmulas-elemento (desde que, é claro, essas formas

tenham sido montadas corretamente – não de um completo improviso supérfluo, como veremos mais a frente);

3. Que esses elementos da expressão presentes na estrutura foram definidos pelas possibilidades combinatórias implicadas por nossas fórmulas sígnicas.

B. O que não temos certeza

1. Onde, quando ou por quem essa língua foi usada, ou se já fora ao menos usada (uma possibilidade a ser considerada é a de que tal língua nunca foi utilizada, mas sim transformada desde o seu início nas direções implicadas pelas diversas línguas sob comparação);
2. Se essa língua contém em sua estrutura mais elementos da expressão do que aqueles designados por nossa fórmulas-elemento;
3. Se os elementos da expressão que sabemos fazer parte da língua foram definidos na estrutura por mais possibilidades combinatórias do que aquelas que são implicadas por nossas fórmulas sígnicas.
4. Como o uso linguístico (se houver um; veja o ponto 1) da língua era organizado e, em particular, (a) que palavras completas ou sentenças existiram, ou poderiam existir, (b) como os elementos da expressão eram representados na fala, na escrita ou por outras maneiras e (c) que significado estava contido nos signos.

Nesses últimos quatro pontos, nos quais representamos nossas incertezas, podemos, se assim o quisermos, construir *hipóteses* baseadas em probabilidades. Não se pode negar que tais hipóteses são extraordinariamente tentadoras e que, em certa medida, elas surgem naturalmente. Aqui temos um terreno fértil para aquelas mentes romanticamente dispostas e imaginativas. Ao combinar-se probabilidade e fantasia, podemos conjurar um amplo quadro de uma língua pré-histórica como a língua originária indo-europeia, com palavras e sentenças, pronúncia e significados bem definidos, todos construídos para “lembrar” o máximo possível palavras, sentenças, pronúncias e significados nas línguas historicamente conhecidas. E, a partir disso, outro quadro pode ser desenhado: toda uma sociedade e cultura pré-histórica, tudo baseado em conjecturas sobre quais palavras existiram e que coisas elas designaram. E, ainda hipoteticamente, a essa sociedade pré-histórica pode ser designado um determinado lugar outro com base em conjunturas, digamos, baseadas nas plantas e animais que por ventura foram designados pelas palavras nessa língua. Naturalmente, a ciência não se absteve de

construir e discutir tais hipóteses, o que é bastante justificável – contanto que se saiba que elas não serão nunca mais do que apenas hipóteses. Entretanto, a linguística clássica estava a caminho de fazer dessa hipótese a parte essencial da comparação genética e, assim, contribuiu para trazê-la ao descrédito injusto de uma ciência insegura e hipotética. Na verdade, é justamente o oposto. É uma ciência com resultados seguros e um método exato. Mas só é possível manter-se assim longe do domínio das hipóteses, apegando-se ao que se sabe com certeza e respeitando os limites de seu método. De forma alguma o método nos permite inferir qualquer informação sobre a cultura, as pessoas ou questões de “raça” de uma língua. O fato de nossa língua ser indo-europeia pode, de fato – se quisermos utilizar uma terminologia inapropriada – ser expressado ao se dizer que nós somos indo-europeus. Porém, não se pode alocar qualquer outro significado que se encaixe melhor nessa sentença do que dizer que nossa língua está conectada às outras línguas indo-europeias por funções-elemento. Isso significa que o sistema de elementos em nossa língua – e pequenas partes dessas palavras – descendem de uma unidade linguística indo-europeia predecessora, mas isso não quer dizer de maneira alguma que nós mesmos, culturalmente, estejamos também conectados a essa unidade. Isso ocorre precisamente porque as línguas podem se espalhar, mover-se e ser transferidas a populações bem distintas daquelas as quais pertenceram originalmente (como ocorreu com o latim, que começou em uma pequena área na nascente do rio Tibre e fira disseminado sobre um enorme número de povos diferentes nos mais variados aspectos, ou como o inglês, mas recentemente, tem se espalhado ao redor do mundo). Portanto, é impossível estabelecer tal conexão entre línguas e povos. Não é possível sequer ter qualquer certeza se uma pequena população que falou uma língua há milhares de anos (como, por exemplo, a pequena comunidade que falava latim na nascente do rio Tibre) realmente possui algum descendente direto atualmente ou, caso possua, se esses descendentes falam uma língua que se conecta geneticamente àquela que fora falada por seus ancestrais. Esse elemento de incerteza obviamente torna-se bem mais amplo conforme ampliamos também o número de famílias linguísticas envolvidas e, conseqüentemente, o quanto voltamos no tempo em nossa datação hipotética das línguas originárias. Portanto, argumentos linguísticos não podem ser usados para dar suporte a quaisquer teorias raciais.

Como vimos, as funções-elemento organizam-se arbitrariamente em relação a maneira como os elementos são representados nas línguas analisadas (seja, por exemplo, por letras ou por sons), e os fórmulas-elemento devem, portanto e igualmente, ter um

caráter abstrato. Com certeza é verdade que os elementos que comparamos nas diferentes línguas indo-europeias, por exemplo, são representadas por letras: afinal de contas, no caso das línguas mortas como o grego antigo, o latim, o hindi antigo e o gótico, não temos outra coisa para trabalhar senão caracteres escritos e a partir disso podemos apenas conjecturar hipóteses mais ou menos prováveis sobre a pronúncia deles. Porém, se não argumentamos a partir das letras dessas línguas que a língua originária do indo-europeu também possuía letras não é meramente porque por acaso sabemos que os alfabetos grego e latino (e hindi) são transformações do alfabeto semítico e, conseqüentemente, possuem uma origem não-indo-europeia. Pelo contrário, é, acima de tudo, porque nosso método simplesmente não nos permite dizer nada sobre como os elementos das línguas originárias eram representados. Pode-se dizer o mesmo sobre (alguns podem ficar tentados a dizer a afirmar isso em um nível até maior) a representação de certos sons na pronúncia. Nesse aspecto, apenas hipóteses podem – e nenhuma hipótese será necessária – ser oferecidas. Tais hipóteses podem ter um nível maior ou menor de probabilidade, mas sempre é uma questão de probabilidade, nunca de certeza. A hipótese de que I.E. **m* fora pronunciado aproximadamente como *m*, em inglês, possui uma alta probabilidade; a probabilidade de que I.E. **A* fora pronunciado aproximadamente como a última vogal em *casa* (uma das hipóteses que fora sugerida) é bem baixa. Portanto, é um infortúnio que a linguística clássica tenha chamado as funções-elemento de *regras sonoras* e, por exemplo, tenha chamado as funções consonantais descobertas por Rask no germânico (p. 41) e no alto germânico (p. 46) de *mudanças sonoras* (Rask, ele mesmo, chamava-as de *mudanças de letras*, o que é consideravelmente uma terminologia bem mais honesta e sóbria).

Da mesma forma, as minúcias do significado das palavras (ou, melhor, as partes das palavras) da língua originária permanecem hipotéticas e, como as minúcias da pronúncia, são supérfluas e irrelevantes do ponto de vista da teoria. Nesse ponto, também, o grau de probabilidade pode ser maior ou menor. As línguas raramente mostram tal acordo em relação ao significado como demonstram em relação a pronúncia de um elemento (como a pronúncia em diversas línguas pode ser conhecida ou determinada em um alto grau de probabilidade); muito raramente encontramos tal unanimidade sobre o significado de uma palavra como encontramos em relação à pronúncia do elemento **m*, por exemplo. Pode-se pensar que existe um alto grau de probabilidade na assunção de que I.E. **patér* signifique “pai”. Mas, em primeiro lugar, **patér* é composto de uma raiz **pA-* e um sufixo derivacional **-t-r*, o qual contém ainda um terceiro componente: a vogal **ē*

que entra em alternância com a vogal * \bar{o} (veja a página 49). A raiz * p_A- está em uma alternância regular com outra forma: * $p\bar{o}-$, a qual aparece em muitas línguas indoeuropeias (por exemplo, em h.a. *pá-ti* “ele protege”) com significados como “proteger” etc. O sufixo *-*t-r* é conhecido a partir de muitas outras palavras. Nas muitas línguas indoeuropeias ele designa a pessoa que exerce uma ação (por exemplo, gr. *rhē-tōr* “orador”, lat. *ōrātor*). Em outras palavras, **pa-tér* é uma formação regular cujo significado podemos supor que tenha sido algo como “o protetor”. Não é necessário que essa formação tenha existido na língua originária; ela é tão regular e natural que poderia ter sido formado independentemente, em tempos antigos, nas diferentes línguas indoeuropeu. Tendo existido ou não na língua originária, não há nada que indique que o significado original tenha sido “pai”.

Uma razão para que o conteúdo do significado continue puramente no campo das hipóteses é que, via de regra, não somos capazes de reconstruir palavras inteiras, mas apenas *signos mínimos*, ou seja, os menores componentes das palavras (tais como raízes, sufixos, e terminações flexionais que podem ser bastante significativas em diversas línguas. A partir do momento em que nos vemos diante de uma formação regular no indoeuropeu, temos que admitir a possibilidade de que ela tenha surgido em qualquer momento, na língua originária ou em uma posterior, assim como as terminações -eiro e -ista, que denotam um agente, em português e que podem ser adicionadas a qualquer substantivo, mesmo que nunca tenha sido antes. Essas formas correspondem ao verbo em dinamarquês *leje* “contratar” (got. *leihwan*), vemos também a forma grega *leípein* “abandonar”. Em grego, no presente, *leípō* é entendido como em I.E. **leik₃-ō*. Em latim, o presente é *linquō*, que equivale ao I.E. **li-n-k₃-ō* (no qual o **e* da raiz alterna com um morfema zero – veja a página 49 – e um componente *n* fora afixado na raiz). No hindi antigo, o presente do mesmo verbo é *riñácmi*, que equivale ao I.E. *li-nē^h-k₃-mi* (com uma terminação diferente, -*mi*, para a primeira pessoa do singular e com uma grade de alternância diferente para o componente infixado). Todas essas formações representam *possíveis* tipos de formações do presente para tal verbo no indo-europeu, mas não podemos afirmar qual deles fora usado na língua originária – possivelmente nenhum, ou talvez um ou mais.

Assim, reconstruímos não palavras, mas partes de palavras sem termos conhecimento de sua pronúncia ou de seu significado. Isso pode parecer soar lamentavelmente pouco – mas, de fato, é um grande empreendimento. Ao usarmos a parca informação a qual temos acesso, podemos provar a existência de relações genéticas em

um espectro linguístico muito amplo e, na verdade, ilimitado. Podemos empreender uma extensa descrição sobre a estrutura de uma língua originária e, por conseguinte, explicar um grande número de detalhes em várias línguas pertencentes a essa família. O método nos permite ainda *prever*. Permite-nos, por exemplo, inferir com certeza se a palavra *mãe* existiu em gótico e, caso não tenha sido submetida aos processos de influência dos elementos, transformação analógica, abreviação ou tabu, sua deve ter sido *modar* – do que podemos ter um alto grau de certeza. Em outras palavras, a reconstrução pode tomar um caminho reverso e, dessa forma, podemos reconstruir a forma que uma palavra pode ter tido nas línguas individuais dentro da família de línguas. Tal reconstrução é, de fato, um dos pré-requisitos para se distinguir entre os empréstimos e as palavras geneticamente relacionadas como também entre as funções-elemento e os contraexemplos.

A linguística genética pode então ser reconhecida como absolutamente exata e é, por ventura, o domínio mais exato dos estudos humanísticos.

Como já havíamos notado, apenas uma pequena fração de palavras em uma língua pode, completamente ou em parte, demonstrar conexões genéticas entre palavras em línguas relacionadas – seja por causa de novas formações, empréstimos ou transformações. Apenas dentro de uma pequena porção do vocabulário é que a linguística genética pode usar o método das funções-elemento. É, portanto, de certa forma impossível para a linguística genética dar conta de todas as palavras de uma língua usando os meios desse método. Mesmo que tentemos sempre descrever o maior número de palavras quanto for possível por meio desse método, sempre respeitando as funções-elemento, a linguística genética pode apenas responder questões *etimológicas*, de forma inequívoca e certa, em uma extensão bastante limitada, ou seja, a história individual de determinadas palavras. Na maioria dos casos, esse é um problema extremamente complicado que não possui nenhuma solução inequívoca e para o qual a linguística genética é inadequada. Um dicionário etimológico toma para si a tentadora, mas não científica, tarefa de contabilizar *todas* as palavras em uma língua. Ao fazer isso, força o método muito além de suas capacidades e traz, com isso, um descrédito que o método não merece. Para estimar a probabilidade das hipóteses etimológicas que tal dicionário oferece para a origem de uma palavra, deve-se conhecer antes as funções-elemento pertencentes a família descrita, de modo que seja possível analisar se as hipóteses concordam com elas. Um leigo, que não tenha estudado as funções-elemento, não pode, de fato, usar um dicionário etimológico e ter um julgamento razoável e justo dessas hipóteses.

Relação tipológica das línguas

Observamos anteriormente que pode ser possível estabelecer funções entre línguas para além das funções-elemento. Adicionalmente ao fato de serem ordenadas em famílias, as línguas podem também ser ordenadas em *tipos* linguísticos de acordo com sua estrutura. Tal classificação depende, é claro, de quais características estruturais são usadas como critério. Assim como o ser humano pode ser classificado em alto, baixo, branco, negro, amarelo, pardo etc. Da mesma maneira, o universo de línguas pode ser dividido de diversas formas, a depender do ponto de vista. Uma possibilidade de escolha é entre a classificação de acordo com a estrutura linguística e a classificação de acordo com o uso linguístico, cada uma com várias possibilidades de subclassificações sob cada caso.

A única tipologia linguística a alcançar um lugar na linguística clássica fora a classificação de acordo com o uso linguístico. O foco central era a estrutura dos signos, especialmente das palavras. *Palavras* são signos permutáveis, signos que podem trocar de lugar com outros signos em uma cadeia linguística: *respondeu suavemente* consiste de duas palavras, porque pode-se dizer também *suavemente respondeu*; *suave-mente* e *responde-u* são formadas por dois signos cada, mas a ordem em que eles aparecem não pode ser outra. Signos permutáveis atraíram uma quantidade extraordinária de atenção por parte da linguística clássica, desde a antiguidade, porque se pensava, em acordo com a lógica conceitual aristotélica, que cada signo representava um conceito. Isso também explica o interesse que existe em se dividir o universo das línguas de acordo com a estrutura de palavras escolhida pelo uso linguístico.

Via de regra, distingue-se quatro tipos linguísticos:

1. Um tipo *isolante*, no qual as palavras são invariáveis, ou seja, não se sujeitam à inflexão ou à derivação e no qual as relações gramaticais podem ser expressas apenas pela ordem das palavras (como em português nas orações *o cachorro mordeu o homem* e *o homem mordeu o cachorro*, nas quais apenas a diferente ordenação das palavras expressa as relações gramaticais demonstradas em outras línguas pelos casos) ou pela adição de palavras independentes (por exemplo, o plural pode ser expresso adicionando-se “muitos” à forma que também serve como singular, ou o presente do indicativo que pode ser assinalado pela palavra “agora”). O exemplo prototípico sempre relacionado a esse tipo é o chinês clássico (o chinês moderno falado classifica-se diferentemente).

2. Um tipo *aglutinante*, no qual todas as relações gramaticais (derivações e inflexões) são expressas por sufixos: signos que são encaixados *depois* do componente lexical ou depois da raiz da palavra. Assim, cada relação possui seu próprio sufixo. Em turco, no qual o sufixo *-lar-* expressa o plural, *-a* o dativo, *-da* o locativo e *-dan* o ablativo, a palavra *kuş* “pássaro” é declinada da seguinte forma:

	singular	plural
nominativo	kuş	kuş-lar
dativo	kuş-a	kuş-lar-a
locativo	kuş-da	kuş-lar-da
ablativo	kuş-dan	kuş-lar-dan

As línguas altaicas e urálicas são os principais exemplos desse tipo.

3. Um tipo *flexional* (a palavra flexão é tomada aqui em seu sentido restrito), no qual a fronteira entre a raiz e o sufixo não é claro. Nesse caso, um mesmo sufixo pode expressar diferentes relações gramaticais ao mesmo tempo (como acontece no latim -*ārum* em *bonārum*; veja a p. 52) e no qual a própria raiz pode mudar durante a flexão com alternâncias vocálicas ou com a inserção de outros componentes.

As antigas línguas indo-europeias servem como modelos para esse tipo.

4. Um tipo *polissintético*, no qual todas as relações gramaticais em uma sentença podem ser expressas por afixos ou por versões alternativas de uma mesma raiz. Uma sentença é normalmente equivalente a uma única palavra e nenhum de seus componentes é permutável. O groenlandês pode ser citado como um exemplo disso: uma sentença como *kavfiliorniarumagaluarpunga* “eu gostaria de fazer um café” contém apenas uma palavra.

Muitas objeções foram feitas a essas classificações. Uma de suas características insatisfatórias é o fato que os tipos quase nunca são encontrados em uma forma pura, justamente porque a maioria das línguas apresenta misturas. Essa classificação fora estabelecida originalmente por Friedrich von Schlegel e Wilhelm von Humboldt. No presente momento, o linguista americano Edward Sapir deu continuidade e desenvolveu essa concepção de uma maneira bastante engenhosa, mas extremamente complicada, em um sistema cujo propósito era fornecer um lugar para todas as estruturas de palavra concebíveis. Nessa forma desenvolvida e esclarecida, a tipologia da estrutura da palavra merece consideração. Não usaremos o elaborado esquema de Sapir aqui, mas

consideraremos apenas brevemente o que pode ser apreendido ao se organizar tais tipos linguísticos.

Apesar da tentativa original concentra-se em uma identificação de tipos linguísticos com determinadas famílias – o tipo isolante com o tai-chinês, o aglutinante com o altaico e o urálico, o flexional com o indo-europeu e o polissintético com certas línguas ameríndias, por exemplo – logo percebeu-se que essa tentativa não obteria sucesso. Seja qual for a configuração dos tipos linguísticos, fica evidente que eles nunca coincidirão com as famílias linguísticas. O próprio fato da mudança linguística, que subentende o surgimento das famílias linguísticas, implica que uma língua pode mudar de tipo ao longo do tempo. Mesmo que se possa sustentar que o indo-europeu tenha sido uma língua flexional, muitas línguas indo-europeias passaram a outros tipos sem, é claro, deixarem de ser indo-europeias. Um esquema declinativo do armênio como o demonstrado abaixo ilustra claramente um tipo aglutinante:

	singular	plural
nominativo	<i>hay</i> “um armeno”	<i>hay-er</i>
dativo-genitivo	<i>hay-u</i>	<i>hay-er-u</i>
instrumental	<i>hay-ov</i>	<i>hay-er-ov</i>
ablativo	<i>hay-ē</i>	<i>hay-er-ē</i>

Fora proposto que essa mudança ocorreu sob influência da língua turca; mesmo que isso seja verdadeiro, o armênio, mesmo em sua forma mais antiga, já não pertencia mais ao tipo flexional.

O inglês moderno, outra língua indo-europeia, tornou-se isolante e, nesse aspecto, aproxima-se mais do chinês do que de sua origem genética. O mesmo pode ser afirmado sobre o francês moderno segundo alguns linguistas, mas na opinião de outros essa língua está indo em direção à polissíntese: na sentença *je ne le lui ai pas donné* “eu não dei a ele”, apenas uma permutação é possível (*ne le lui ai-je pas donné?* “eu não dei a ele?”), o que nos faz lembrar o tipo de frase-palavra que encontramos em groenlandês.

Assim, as relações genéticas e tipológicas são duas coisas bem diferentes e não têm nada a ver uma com a outra. Dentro de uma mesma família, podemos encontrar línguas com tipos fundamentalmente diferentes e dentro de um mesmo tipo, encontrar línguas de famílias distintas. Essas duas divisões do universo de línguas não possuem nada em comum. Elas são estabelecidas sob bases diferentes e a partir de pontos de vista

completamente diferentes. Da mesma forma, o zoologista pode classificar sistematicamente os animais como mamíferos, peixes etc. mas também ecologicamente organizando, por exemplo, uma categoria de animais marinhos que incluiria baleias, focas, polvos, ostras e peixes – animais que pertencem a categorias bem distintas em uma classificação sistemática.

A linguística clássica esteve bem menos interessada nas relações tipológicas do que nas relações genéticas entre línguas. Mas não é necessária uma longa reflexão para perceber que esse é um campo muito rico e importante, mesmo que ainda quase totalmente deixado de lado. É claramente possível organizar uma tipologia linguística a partir de muitos outros pontos de vista para além desse que foi arbitrariamente escolhido pela linguística clássica e, como observamos, demonstra ser o mais superficial de todos – aquele que considera a estrutura da palavra. Também é fácil de ver que, independentemente do ponto de vista escolhido, a situação permanecerá a mesma: a classificação resultante passará despercebida pela classificação genética e não terá qualquer relação com ela.

Quando tentamos organizar uma tipologia linguística a partir de pontos de vista mais profundos e imanentes do que os escolhidos até o momento, logo veremos que as entidades que devem ser comparadas nas diferentes línguas são *categorias*. Essas serão (1) categorias na estrutura linguística (como vogais e consoantes, acentos e modulações e suas várias subcategorias no lado da expressão e categorias gramaticais como o caso, gênero e número no lado do conteúdo) e (2) categorias no uso linguístico (como as categorias da expressão – sons – e as categorias do conteúdo – significado). A partir de todos esses pontos de vista, as línguas que demonstrarem as mesmas categorias serão alocadas no mesmo tipo linguístico. A comparação entre elementos, por outro lado, nunca será de interesse para a linguística tipológica, na qual o fato de que duas línguas possuem um elemento em comum sempre significará que elas possuem uma categoria em comum – a categoria representada por aquele elemento. Isso ficará mais claro a partir do que veremos a seguir, mas já estamos em uma posição em que podemos declarar de que esse é motivo pelo qual as relações genéticas e as relações tipológicas são, essencialmente, coisas diferentes. Ao termos definido operacionalmente a relação genética como uma função entre línguas formada por elementos-função, podemos dar uma classificação correspondente para *a definição operacional da relação tipológica como uma função entre línguas formada pelo fato de que categorias em cada língua têm uma função com categorias em cada uma das outras línguas*. Assim como a relação genética é reconhecida

a partir das funções-elemento, a relação tipológica, quando vai além do nível mais superficial e acidental (o da estrutura da palavra), é reconhecida a partir das *funções categoriais*.

Tal tipologia também levará em conta, naturalmente, a estrutura das unidades (sílabas, por exemplo) que estão na língua, uma vez que a estrutura dessas unidades depende simplesmente das categorias encontradas na língua. Definimos uma categoria como a soma das entidades que podem ser inseridas em determinados lugares na cadeia (p. 54) e demonstramos ainda que as vogais e as consoantes são exemplos de categorias (p. 56). Uma unidade como a sílaba, então, dependerá nas categorias de elementos existentes e assim também para todas as outras unidades.

Uma tipologia linguística exaustiva é, de fato, a maior e mais importante tarefa da linguística. Diferente da linguística genética, ela não possui limitações regionais. Seu objetivo final deve ser demonstrar quais estruturas linguísticas são possíveis, em geral, e porque somente essas estruturas são possíveis e não outras. E aqui ela terá se aproximado mais do que qualquer outro tipo de linguística do que poderia ser chamado como o problema da essência da língua. Por fim, ela demonstrará ser superordenada à linguística genética, pois apenas por meio da linguística tipológica é que poderemos entender que leis governam a mudança linguística e que possibilidades de mudança um determinado tipo linguístico implica. Apenas por meio da tipologia é que a linguística pode ascender aos pontos de vista mais gerais e tornar-se uma ciência. Por ora, com certeza, muito pouco foi realizado dessa grande proposição. A tarefa está lançada, mas está longe de ser completada. A linguística clássica esteve seriamente interessada apenas no uso e na genética, enquanto a estrutura e a tipologia foram deixadas quase completamente de lado. Os problemas centrais não foram resolvidos – não foram ao menos colocados à mesa – pela linguística clássica. Isso é deplorável, se me permite, e o leitor pode também deplorar o fato de que a exposição a seguir da linguística tipológica necessariamente apenas consistirá de não mais que sugestões incompletas – um programa incompleto, um questionário com espaços vazios. Mas para o estudioso não há nada mais belo do que a visão de uma ciência por criar-se.

Tipologia das estruturas linguísticas

Como uma categoria é formada por entidades que podem ocupar um determinado lugar na cadeia, devemos começar pela cadeia como o dado imediato para determinar a estrutura de qualquer língua em particular (état de langue); é por meio de uma análise da cadeia que chegaremos às categorias. Uma cadeia linguística inteira que é assim tomada como um objeto é chamada de *texto*.

Naturalmente, a análise não deve ser um recorte ao acaso do texto, mas deve ser feita levando em conta as *relações* entre suas partes constituintes (veja a p. 53). Essas relações podem ser de diversos tipos.

Primeiro, temos de distinguir entre *relações implicativas* e *relações não-implicativas*. Diremos que duas entidades possuem uma *relação implicativa* sempre que a presença de uma na cadeia for uma pressuposição necessária para a presença da outra ou, dito de outra forma, sempre que uma das entidades necessariamente implicar a existência da outra. E diremos que existe uma *relação não-implicativa* quando não houver tal pressuposição entre os elementos relacionados – quando podemos ter qualquer uma das entidades sem necessariamente precisarmos da outra.

Relações não-implicativas são muito comuns entre elementos. Na palavra *blusa* existe uma relação entre *b* e *l*: eles podem aparecer somente no campo inicial da sílaba. Mas a relação é do tipo não-implicativa; encontramos *l* sem *b* (*lusa*) e *b* sem *l* (*busa*) – do verbo *busar* “soprar”, em português angolano. Via de regra, podemos também encontrar tal relação não-implicativa desde que consideremos apenas uma consoante em particular em sua relação com alguma vogal em particular. Existe uma relação entre *b* e *e* na sílaba *be*, mas a relação é não-implicativa: podemos ter uma sílaba com *b* e sem *e* (*ba*) e uma sílaba com *e* e sem *b* (*me*). No entanto, acontece que – depois de termos estabelecido as categorias com base na análise textual – existirá uma relação implicativa entre a categoria das vogais e a categoria das consoantes em uma sílaba: não podemos ter um representante da categoria das consoantes sem ter também um representante da categoria das vogais na mesma sílaba (veja a p. 55). Da mesma forma, existe uma relação não-implicativa entre uma oração subordinada específica e uma oração principal específica, porque a oração subordinada pode ser substituída por outra oração subordinada e a oração principal por outra oração principal; mas existe uma relação implicativa entre a categoria das orações subordinadas e a categoria das orações principais. Entretanto, é possível encontrar

relações implicativas entre elementos específicos e, como veremos a seguir, existem muitos exemplos disso, tanto no nosso dia a dia como no mundo da linguagem.

Relações implicativas podem ser posteriormente classificadas como *recíprocas* ou *unilaterais*. Existirá uma relação implicativa recíproca entre duas entidades quando ambas se pressupõem mutuamente; quando uma pressupõe a outra, mas o inverso não é verdadeiro, a relação é unilateral. Podemos citar como exemplos de uma *relação implicativa recíproca* a relação entre o volume um e o volume dois de um livro, entre parceiros em um casamento ou entre dois noivos – eles não poderiam receber esse título a não ser que os dois o sejam. Vamos aprofundar nosso raciocínio: se um autor escreve um livro e o chama de volume um, ele pressupõe, por chamá-lo assim, a existência do volume dois, mesmo que esse segundo volume nunca seja publicado. Da mesma forma, se um homem é chamado de marido, a existência de sua esposa é pressuposta, mesmo que nunca venhamos a conhecê-la. Em outras palavras, podemos nos deparar com apenas uma entidade em uma relação implicativa e, baseando-nos nisso, poderemos *suprir* a relação não com qualquer outra entidade concreta em particular que poderíamos descrever em detalhe, mas com uma entidade sobre a qual sabemos apenas ser algo pressuposto pela primeira. Em tal caso, em que uma entidade pressuposta não está manifestada, mas pode ser suprida, diremos que é uma entidade *latente*.

Se uma língua é construída de modo a não aceitar sílabas que contenham uma vogal sozinha, mas apenas sílabas que contenham ambas vogais e consoantes, diremos que existe uma relação implicativa recíproca entre a categoria das consoantes e a categoria das vogais nessa língua. Mas se as sílabas de uma língua consistirem de uma vogal sozinha ou de uma vogal acompanhada por uma consoante, podemos concluir que essa língua demonstra uma *relação implicativa unilateral* entre a categoria das consoantes e a categoria das vogais: a consoante *implica* a vogal, mas não o contrário; a vogal é *implicada* pela consoante, mas não o contrário. A relação implicativa unilateral é a relação mais comum entre a categoria orações subordinadas e a categoria de orações principais: a oração subordinada implica a oração principal, mas não o inverso; a oração principal é implicada pela oração subordinada, mas não o inverso. E caso encontremos uma oração subordinada sozinha sem nenhum traço de sua oração principal, concluiremos que a oração principal é latente e suprimo-la conforme necessário – não com uma oração concreta em particular, mas uma sobre a qual sabemos apenas que é a oração principal implicada pela oração subordinada em questão. Uma exclamação como *Ah se eu tivesse dinheiro!* requer uma suplementação desse tipo. Exemplos de relações implicativas

unilaterais são, é claro, fáceis de encontrar fora da língua também. No xadrez, o rei é implicado por todas as outras peças – se o rei está em xeque, todos os seus subordinados são capturados e o jogo acaba – mas o contrário não é verdadeiro.

Para nos certificarmos de que executamos uma descrição exaustiva da língua, de tal forma que todas as relações sejam registradas, não podemos diretamente quebrar o texto em seus menores elementos que possuem relações demonstráveis, mas devemos tornar o texto objeto de uma série de sucessivas divisões de tal forma que cada passo terá o menor número de partes que for possível. Dessa forma, qualquer texto linguístico poderá ser dividido primariamente em duas partes, um *plano do conteúdo* e um *plano da expressão* que possuem uma relação implicativa recíproca: se estamos tratando realmente de uma língua, ela deve conter tanto um plano do conteúdo como um da expressão. Cada um desses planos é então dividido novamente em partes e em partes de partes – capítulos, parágrafos, sentenças, orações etc. – até os menores elementos.

A cada passo dessa divisão, podemos organizar categorias por alguma das relações que viemos discutindo até o momento. Uma categoria – por exemplo, das orações subordinadas ou das consoantes – será definida como implicativa; outra – digamos que a categoria das orações principais ou das vogais – será definida como implicada; e assim por diante. Um fato interessante é que os membros de cada categoria possuem a propriedade peculiar em que a substituição de um membro por outro pode acarretar uma mudança no plano oposto da língua. Os elementos da expressão, por exemplo, que são membros das categorias de consoantes e de vogais serão, cada um, definidos como distintos dos outros membros das mesmas categorias pelo fato de que a substituição de qualquer um desses por outros acarretará em uma mudança no conteúdo linguístico (no significado de um signo). O fato de que *p*, *s*, *r* e *t* são quatro elementos da expressão em português e que *a*, *e*, *i* e *o* são quatro elementos da expressão também distintos em português pode ser definido de imediato pois, se trocarmos qualquer um por outro no mesmo lugar da cadeira, observaremos uma mudança no plano do conteúdo: *pat*, *sat*, *rat*, *tat*, *pat*, *pet*, *pit*, *pot* (veja a p. 54). Somos capazes de determinar o número de membros em uma categoria linguística a partir do que chamamos de *teste de comutação*. Obviamente, isso também é válido para os elementos do conteúdo da língua: “nominativo” e “genitivo”, “presente” e “pretérito”, bem como todos os componentes lexicais, são linguisticamente diferentes pelo fato que a mudança de um acarretará uma alteração no plano da expressão. Também fica claro que esse teste também é aplicável não somente ao que chamamos aqui de elementos da expressão e elementos do conteúdo,

mas a todos os membros, sem exceção, de quaisquer categorias. Por exemplo, duas sentenças são diferentes apenas se seus conteúdos são diferentes e se suas expressões são diferentes. Se mudarmos meramente a expressão de uma sentença, de forma que isso não altere o seu conteúdo, produzimos apenas uma mudança na pronúncia, não uma diferença linguística; e caso mudemos somente o conteúdo de uma sentença, de forma que isso não altere sua expressão, produzimos apenas uma mudança de pensamento, não uma diferença linguística.

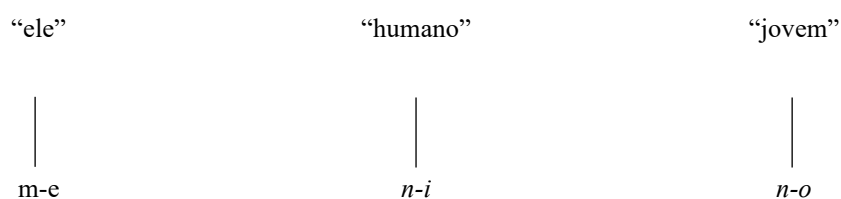
Assim, chegamos a uma propriedade fundamental da língua, aquela que a distingue de outras estruturas. A partir da palavra *comutação*, no sentido especial em que a temos usado, podemos dar a definição concisa de que *a língua é uma estrutura na qual os membros de cada categoria possuem mútua comutação*.

Naturalmente, essa definição é bem ampla, por isso, quando a aplicamos, devemos estar preparados para descobrir que ela é satisfeita por muitas estruturas que talvez não estejamos acostumados a chamar de línguas, pelo menos não línguas comuns do dia a dia. Então, por exemplo, veremos que tudo que chamamos de uniformes, no sentido mais amplo, incluindo aqui trajes folclóricos e semelhantes, constituirá um sistema que segue essa definição. Todo uniforme possui um conteúdo correspondente: esse uniforme “significa” um oficial francês de certa patente; essa vestimenta “significa” um prelado católico romano de certo nível. Quando, como costumeiramente acontece, existem diferentes vestimentas para jovens garotas e para mulheres casadas, cada uma desses “uniformes” possuirão “significados” especiais. Podemos, portanto, aplicar aqui o teste de comutação: diferentes conteúdos correspondem a diferentes expressões (diferentes vestimentas) e vice versa. Teremos assim um considerável número de outras áreas em que encontraremos estruturas que provam ser, em relação a esse aspecto decisivo da comutação, idênticas às estruturas que estamos acostumados a chamar de línguas. Na medida em que eles tentam comunicar um conteúdo, os vários gêneros artísticos como a pintura descritiva, música programática e assim por diante, devem ser considerados línguas segundo a nossa definição. Um jogo, se interpretado, também será uma língua nesse sentido – como, por exemplo, um jogo de xadrez se interpretado como duas matrizes de batalha, cada uma formada por oficiais e soldados de maior e menor patente e, dentre os oficiais de maior patente, de um rei e uma rainha e assim por diante. Em tal caso, temos ambos conteúdo e expressão e, portanto, podemos aplicar o teste de comutação.

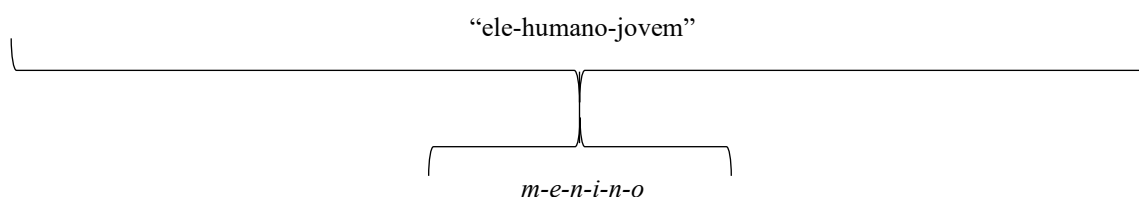
No entanto, pode ser útil afunilarmos nossa definição ao adicionarmos a seguinte propriedade: se olharmos para as línguas comuns do nosso dia a dia, deveremos encontrar em sua estrutura a propriedade particular de que um único elemento da expressão não costuma, via de regra, corresponder a (ou seja, não possui relação) um único elemento do conteúdo e vice versa. Como vimos anteriormente (p. 53), *pode* acontecer de um elemento da expressão corresponder a um elemento do conteúdo (-s como expressão do genitivo em inglês, por exemplo), mas esse é um caso especial e não corresponde a toda a extensão da estrutura da língua. Via de regra, um signo é composto de dois ou mais elementos da expressão e dois ou mais elementos do conteúdo conectados a eles. Agora, no que diz respeito à expressão linguística, somos felizes em saber— ou imaginar que sabemos — quantos elementos da expressão estão em um determinado signo. Já em relação ao conteúdo linguístico, no presente momento da pesquisa, ainda não temos informações suficientes. Porém, podemos facilmente executar uma análise provisória que será adequada ao nosso propósito no momento. O plano da expressão das palavras *touro* e *vaca* podem ser analisadas da seguinte forma: a primeira possui as letras *t, o, u, r* e *o* enquanto que a segunda possui *v, a, c* e *a*. Já em relação ao plano do conteúdo cada palavra pode ser analisada em dois elementos: “ele-bovino” e “ela-bovino”, respectivamente. Poderíamos certamente avançar mais no procedimento ao analisarmos a palavra *bovino* (e, talvez, as palavras *ele* e *ela*) em componentes, mas isso nos levaria a frente do nosso propósito atual. Como já pode ser percebido, o que fizemos na análise do conteúdo de *touro* e *vaca* foi substituir o signo-conteúdo simples por um conteúdo composto cujos componentes fazem parte de outros signos-conteúdo na língua. Assim, os componentes “ele” e “ela” fazem parte dos signos *ele* e *ela, menino* e *menina, cavalo* e *égua*. Não questionaremos aqui até que ponto o componente “bovino” entra em qualquer signo-conteúdo composto, até porque a resposta para isso requer a análise de “bovino”, a qual conscientemente não procederemos; de qualquer forma, componente “bovino” obviamente entra no signo-conteúdo da palavra *bovino*. Essa substituição de um signo-conteúdo simples por um que seja composto não envolve qualquer mudança na expressão: é a mesma coisa se dizemos que *touro* significa “touro” ou que significa “ele-bovino”. O que executamos aqui com o plano do conteúdo corresponde completamente ao que estamos acostumados a fazer no plano da expressão. Quando analisamos *touro* em *t, o, u, r* e *o*, estamos substituindo o signo-expressão simples por um composto que é formado por componentes que fazem parte de outros signos-expressão da língua. Reforçamos novamente: essa substituição de um signo-expressão simples por composto não envolve

qualquer mudança no conteúdo: é a mesma coisa se dissermos em português *touro* ou *t, o, u, r e o*. Em ambos os casos, tanto no conteúdo como na expressão, *definimos* a entidade não analisada reunindo outras entidades tomadas da mesma língua. Essa é sempre a maneira pela qual uma definição é feita e as definições de significado que são encontradas em dicionários devem ser entendidas como tentativas de executar o tipo de análise de conteúdo que delineamos. Para evitarmos qualquer mal-entendido, devemos também observar que essa definição não consiste de enumeração desordenada dos componentes dos signos, mas postula um tipo especial de relação entre eles. A definição do plano da expressão da palavra *touro* deve incluir a informação de que *t, o, u, r e o* aparece nessa exata ordem e, analogamente, a definição do plano do conteúdo postula uma certa relação lógica entre os componentes na qual o signo-contéudo é analisado.

Portanto, a língua não é construída de forma que cada elemento terá apenas um elemento da expressão correspondente e vice versa, mas de forma que uma unidade de elementos do conteúdo corresponderá a uma unidade de elementos da expressão e vice versa. A situação que encontramos não é a seguinte:



Mas sim:



Agora poderemos levar em consideração essa propriedade particular em nossa definição de língua ao adicionarmos a seguinte restrição: *se uma estrutura que possui relação implicativa recíproca entre o plano do conteúdo e o plano da expressão deve ser reconhecida como uma língua, é necessário que não exista uma relação implicativa recíproca de um-para-um entre seus elementos de expressão e seus elementos de conteúdo.*

Se adicionarmos essa restrição, não há dúvida de que alguns dos exemplos dados anteriormente não mais serão qualificados como línguas. Essa é uma questão de menor

importância, a qual não desenvolveremos para além disso. Porém, essa restrição é útil por outra razão: fora do domínio das línguas do dia a dia, muitas vezes nos deparamos com a dificuldade de decidir se é necessário dividir uma dada estrutura em um plano de conteúdo e um plano de expressão; porém, com a restrição que postulamos acima, tal divisão deve ser omitida caso exista uma correspondência de um-para-um entre os dois planos de tal forma que dentro de cada plano os elementos sejam definidos precisamente pelas mesmas relações.

Se adicionarmos ou não essa restrição, devemos estar preparados para perceber que várias outras estruturas para além das línguas cotidianas terão qualificativos para serem chamadas de línguas de acordo com a nossa definição. Uma língua cotidiana (como o dinamarquês, o inglês ou o português) é um tipo especial de língua: *por língua cotidiana, deve-se entender uma língua na qual todas as outras linguagens possam ser traduzidas*. Qualquer jogo de xadrez pode ser traduzido – ou explicado – em uma língua cotidiana, mas o contrário não é verdadeiro. Em geral, uma língua cotidiana difere de todos os outros tipos de linguagens (por exemplo, a linguagem simbólica do matemático ou a linguagem de fórmulas do químico) por não ter sido feita especialmente para um propósito em particular, mas sim para servir a todos os propósitos. Em uma língua cotidiana, podemos, com a ajuda de circunlocuções, se necessário, ou explicações minuciosamente preparadas – explicar o que quisermos. Mesmo qualquer peça de música programática poderá ser traduzida em uma peça escrita em uma língua cotidiana – mas não o contrário. Em uma língua cotidiana, como Søren Kierkegaard nos diz, pode-se “trabalhar sobre o inexprimível até que ele seja expresso”. Essa é a vantagem da língua cotidiana – e também o seu mistério. Também é por isso que o lógico polonês Tarski (que chegou a mesma conclusão de Kierkegaard independentemente dele) aponta de forma certa que as línguas cotidianas são caracterizadas em contraste com as outras linguagens por conta de “universalismo”.

Para estabelecer uma teoria geral da estrutura linguística, devemos começar com uma definição de língua a partir do tipo que viemos considerando até agora, porque estabelecer uma tal teoria não é um procedimento simples ou, principalmente, uma questão de experiência, mas sim de cálculo. A experiência não constitui uma base adequada para a teoria da estrutura linguística: seria impossível percorrer todos os textos existentes e, ademais, seria fútil assim proceder, porque a teoria deve ser válida não somente para os textos que foram escritos ou falados até agora, mas também para aqueles que serão escritos e falados no futuro – ou seja, deve ser adequada a todos os textos

teoricamente possíveis e para todas as línguas teoricamente possíveis. Com base em certas experiências, as quais devem necessariamente ser limitadas, mesmo que sejam as mais variadas quanto possível, podemos proceder a um cálculo de todas as possibilidades concebíveis dentro de certos limites. Esses limites são estabelecidos arbitrariamente ao tomarmos – como acabamos de fazer – certas propriedades que estão presentes em todas os objetos que são comumente chamados de línguas e generalizá-las de forma a afirmar: todas os objetos que possuírem essas características serão chamadas de línguas. Pode acontecer que alguns dos objetos que se enquadram nessa definição não sejam comumente chamados de línguas, mas isso não faz diferença desde que não se exclua nenhum objeto que *seja* comumente chamado de língua. Depois que tenhamos definido – arbitrariamente, mas adequadamente – o que devemos entender por língua, procederemos então a um cálculo geral para todos os objetos que provarão satisfazer nossa definição, um cálculo que forneça todos os casos concebíveis. Esse cálculo deriva de nossa definição e funda-se sobre si mesmo, não na experiência, mas apenas pelas regras da lógica. Não devemos nos preocupar se o cálculo será adequado para certas línguas ou certos textos. A teoria da estrutura linguística deve ser estabelecida de forma a possuir consistência e exaustividade de descrição incorporada ao cálculo e esse requisito não pode ser comprovado verificando se o cálculo será adequado a todos os objetos existentes (tal verificação seria impraticável, tanto pelo número e pela extensão dos objetos, mas também porque a teoria deve ser aplicável também a todos os objetos que ainda não existem), mas apenas pelo reexame do cálculo teórico para verificar se ele permanece consistente e exaustivo. Sobre esse aspecto, o teórico da linguística comporta-se como qualquer outro teórico – como um matemático, por exemplo, que organiza sua teoria sem levar em conta aplicações práticas, mas apenas sua teoria; e, precisamente por essa razão, que a teoria pode ser aplicada a casos que ainda não existiam quando ela fora estabelecida. Um engenheiro pode construir uma ponte simplesmente ao ater-se as fórmulas matemáticas. Desde que a teoria do matemático seja consistente e exaustiva, e o engenheiro a conheça suficientemente bem e a aplique apropriadamente, a ponte persistirá.

Um trabalho de cálculo como o que viemos delineando é, obviamente, um esforço de grande envergadura e grande complexidade. Nada parecido fora empreendido pela linguística clássica que, na verdade, nem mesmo vislumbrou essa possibilidade. Muito esforço fora depreendido nessa tarefa e muitas tentativas foram feitas para leva-la a cabo, mas ainda há um extenso caminho a ser percorrido para o estabelecimento de uma

tipologia em que as línguas cotidianas poderão ser distribuídas e, em cuja base, poderá se firmar proposições empíricas gerais sobre a estrutura dessas línguas e suas funções internas. Por essa razão, tudo o que falarmos sobre essa discussão ostentará um caráter provisório.

Podemos considerar como certo que toda língua possui dois, e apenas dois, planos: o plano do conteúdo e o plano da expressão. Na teoria, devemos também considerar a possibilidade de que a língua possa vir a ter, no primeiro estágio de análise, três ou mais planos, mas seria absurdo supor que tal estruturação seja encontrada em qualquer língua cotidiana, portanto, podemos ignorar essa possibilidade em nossa discussão atual. Cada um dos planos, como já vimos, podem ser sucessivamente divididos até as entidades mínimas, as quais chamamos nesse livro de *elementos* – elementos do conteúdo e elementos da expressão. Esses elementos são ordenados em *categorias*. Uma tipologia linguística deve, então, ser capaz de demonstrar quais categorias podem aparecer em uma língua, quais categorias podem aparecer juntas, permitindo ou implicando a presença de outra, e quais categorias não podem aparecer juntas porque uma implica a ausência da outra. Assim como distinguimos, na cadeia, entre relações implicativas recíprocas ou unilaterais e relações não-implicativas; aqui distinguiremos entre *correlações implicativas recíprocas ou unilaterais* e *correlações não-implicativas*.

Em cada um dos dois planos de uma língua, as categorias mais englobantes em que podemos alocar os elementos são as que chamaremos de *elementos básicos* e *elementos caracterizantes*. Essa divisão baseia-se em um tipo especial de relação chamado *regência*. Devemos deixar claro que nem todo tipo de regência será considerada aqui, mas apenas aquelas que chamamos de *diretiva* – o tipo de regência que serve para estabelecer uma oração ou uma unidade composta por orações. Na oração *sa première femme aimait les fleurs* existe um exemplo de regência porque *sa* e *première* estão no feminino (em contraste com *son* e *premier*, que estão no masculino), porque *femme* é feminino e requer que seus adjetivos concordem consigo; mas nesse caso a regência não pode ser caracterizada como uma diretiva porque não estabelece uma oração, mas apenas parte de uma (tal parte, em nosso exemplo, *sa première femme*, é chamada de *junção*). Por outro lado, na oração *sa première femme était jolie* a regência que conecta o substantivo feminino *femme* e a adjetivo feminino *jolie* é uma diretiva já que por meio dele (e de outras regências semelhantes) a oração é estabelecida como uma unidade linguística formal.

Já por *elementos caracterizantes*, devemos entender ou como um elemento que pode entrar como governado em uma regência ou como um elemento que entra em um paradigma acompanhando outro elemento com essas características. No plano do conteúdo da língua, os elementos caracterizantes são os *elementos flexionais* (o gênero “feminino” nos exemplos acima, por exemplo), e os elementos básicos são chamados de *elementos rádicos* (os elementos que entram em palavras lexicais como “touro” e “vaca”). No plano da expressão, os elementos caracterizantes são os *acentos* e as *modulações*. Em dinamarquês existem dois acentos, um manifestado na pronúncia por uma acentuação (relativamente) forte e outra marcada por uma acentuação (relativamente) fraca. Ambas possuem comutação mútua (*kórset* “a cruz” e *korsét* “espartilho”, cujas pronúncias diferenciam-se apenas pela acentuação, possuem conteúdos diferentes) e também uma relação implicativa unilateral: a presença de uma acentuação (relativamente) fraca em uma enunciação pressupõe a presença de uma acentuação (relativamente) forte, mas o contrário não é verdadeiro. Além do mais, essa relação implicativa é uma regência suficiente, por si mesma, para estabelecer uma unidade propositiva da expressão.

Nossos resultados provisórios indicam que toda língua possui em seu plano do conteúdo elementos flexionais (no sentido empregado aqui) e elementos rádicos; mesmo as chamadas línguas isolantes possuem certas “partículas” que entram aqui como elementos regidos em uma diretiva. O plano da expressão, por outro lado, nem sempre demonstra a distinção correspondente entre acento e modulação de um lado e elementos básicos de outro. No dinamarquês escrito no dia a dia não temos nem acento nem modulação no sentido usual dos termos. Certamente, temos sinais de pontuação como o ponto de interrogação, que sinaliza “para a frente” em direção a uma resposta que virá a seguir (que pode, se necessário, ser pressuposta), mas tais marcas de pontuação não constituem uma absoluta necessidade. Muitas línguas não apresentam acentos na pronúncia. O francês é um exemplo disso. Com certeza é possível intercambiar entre uma acentuação mais forte e outra mais fraca na pronúncia do francês, mas nunca é possível distinguir o conteúdo de duas palavras por meio de uma distribuição diferente da tonicidade do acento como vimos em dinamarquês com as palavras *kórset* – *korsét*. Se todas as línguas possuem modulações é uma questão ainda em aberto, até porque a descrição clássica das línguas não deu nenhuma atenção às modulações.

Assim, podemos esperar encontrar no sistema da linguagem humana (no que concerne as línguas cotidianas, que são a base de nossa discussão) uma correlação implicativa unilateral – no plano da expressão, em qualquer caso – entre os elementos

caracterizantes e os elementos básicos: uma língua não pode ter acento e modulação sem ter também os elementos básicos da expressão (vogais e consoantes), mas pode ter estes sem ter aqueles. No plano do conteúdo, pode haver por ventura uma correlação implicativa recíproca entre os elementos flexionais e os elementos rádicos.

Essas duas categorias principais, ou *espécies*, de elementos – caracterizantes e básicos – poder ser posteriormente divididos em *tipos*. Seria muito complicado explicar aqui como essa divisão das espécies em tipos ocorre de acordo com a teoria, mas, na prática, podemos dizer que dividimos os elementos caracterizantes em dois grupos: aqueles que podem ser usados para caracterizar enunciados inteiros ou algo semelhante que entre em um paradigma com estes e aqueles que não possuem essa propriedade. Os primeiros são chamados de *extensos*, enquanto que os últimos são chamados de *intensos*. No plano da expressão, as modulações são os elementos extensos e os acentos são os elementos intensos. No plano do conteúdo, os elementos extensos são aqueles comumente chamados de elementos de flexão verbal (tempo e modo, por exemplo), enquanto que os elementos intensos são aqueles conhecidos como elementos de flexão nominal (caso e número, por exemplo). Os elementos básicos são divididos em *centrais* – aqueles que são indispensáveis em uma unidade mínima, ou semelhante que entre em um paradigma com este – e *periféricos* – aqueles que são dispensáveis em uma unidade mínima. No plano da expressão, no qual a unidade mínima é a sílaba, os elementos indispensáveis são as *vogais* e os elementos dispensáveis são as *consoantes*. No plano do conteúdo encontramos também “sílabas do conteúdo” (elas podem ser chamadas de sintagmas mínimos – ou morfemas – e, na prática, normalmente coincidiram com palavras individuais), divisíveis em elementos indispensáveis ou *elementos radicais* (o elemento “feliz” presente no exemplo “in-feliz-mente”) e os *elementos derivacionais* (os elementos “in” e “mente”).

Após a exposição acima, podemos contar com a possibilidade de que uma língua possa ter modulações sem ter acentos (francês, por exemplo), mas raramente o inverso é verdadeiro. Assim, temos uma correlação implicativa unilateral. É um fato que uma língua não pode ter consoantes se não tiver vogais também, mas isso é simplesmente uma consequência da forma como definimos vogais e consoantes. Em vez de vogais e consoantes, uma língua pode ter uma terceira categoria indiferenciada de elementos da expressão. Dessa forma, existem línguas que possuem uma “harmonia vocálica” (finlandês, húngaro e turco, por exemplo), nas quais um vocoide (som vocálico) em uma cadeia determina a escolha de todos os outros vocoides na mesma cadeia. Em tais línguas, os vocoides representam os acentos, visto que eles entram como elementos governados

(pressupostos) em uma diretiva. E, a partir disso, temos que os contoides (sons consonantais) não representam consoantes nessas línguas, mas sim uma categoria indiferenciada de elementos da expressão. Já no plano do conteúdo, parece ser verdadeiro que uma língua possa ter elementos verbais sem possuir elementos nominais e flexionais (esse é provavelmente o caso em línguas verdadeiramente isolantes, como o chinês clássico), mas o contrário, não. Temos aqui, então, o que parece ser uma correlação implicativa unilateral.

Os tipos, por sua vez, podem ser posteriormente divididos em *subtipos* de acordo com suas relações implicativas. Já para os elementos básicos, o tipo central pode ser dividido em dois grupos: aqueles que devem necessariamente estar presentes na parte central da sílaba (ou sintagma) e aqueles que não precisam estar. Nesse sentido, o tipo periférico pode ser também dividido em dois grupos: aqueles que devem necessariamente estar presentes na parte periférica da sílaba e aqueles que não precisam estar. No plano da expressão, por exemplo, as consoantes podem ser divididas em duas categorias, uma que é representada por sons ou letras que aparecem no começo da sílaba e a outra que é formada por aqueles que aparecem no final – ou uma categoria pode ser representada por aqueles que ocupam o primeiro lugar em um grupo, enquanto que a outra é representada por aqueles que ocupam o segundo lugar. A divisão dos elementos flexionais em subtipos é particularmente importante. Eles podem ser divididos em elementos que podem estabelecer uma sentença composta (ou elementos que entram em um paradigma com ela) e elementos que não podem estabelecer tais sentenças. O caso, por exemplo, poder servir apenas para estabelecer apenas uma proposição: as relações que o caso depreende nunca vão além dos limites de uma proposição. Outras categorias flexionais, entretanto, podem entrar em uma diretiva que vai além dos limites de uma proposição: em *les fleurs sont jolies, et elles sont fraîches* a diretiva entre as formas femininas no plural é observada nas duas proposições; mas o plural feminino *elles* na segunda oração é escolhida por causa do plural feminino presente na oração anterior, a qual ela pressupõe e assim surge ainda outra diretiva no plural feminino unindo as duas orações em um enunciado composto. Esse exemplo ilustra como as diversas “categorias gramaticais” (caso, gênero, tempo, modo etc.) são classificadas na estrutura linguística.

A divisão em subtipos não parece ser uma característica universal a todas as línguas, mas é muito comum na medida em que os elementos flexionais são considerados. Com a ajuda do teste de comutação, podemos descobrir quantos elementos são encontrados em cada uma das categorias em uma determinada língua e que relações

numéricas governam a estrutura da língua humana, em geral, a esse respeito. Tipicamente, o número de elementos é sempre relativamente pequeno. Afinal, a possibilidade de forma um número praticamente ilimitado de signos a partir de um pequeno número de elementos é implicada na própria essência e no propósito de uma língua. Mas dentro das categorias individuais que são representadas em qualquer língua, esse número varia consideravelmente. Existem línguas com duas vogais e línguas com doze, ou quiçá ainda mais, e todas as possibilidades entre essas quantidades. Há, geralmente, mais consoantes que vogais. O número de acentos e modulações sempre parece ser relativamente pequeno, enquanto que no plano do conteúdo o número de elementos flexionais é relativamente pequeno, como em inglês ou dinamarquês, ou bem amplo. Há línguas com dois casos e línguas com mais de cinquenta. Há línguas com dois números (singular e plural), com três (como o lituano ou o grego antigo, que também possui o chamado dual), e com quatro ou cinco (como certas línguas do Pacífico com seus números triádicos e quadrais). Há línguas com dois gêneros (como o português, francês e o dinamarquês, se não levarmos o pronome em consideração), com três (alemão), com quatro (os pronomes dinamarqueses *han* “ele”, *hun* “ela”, *den* “aquilo [gênero comum]”, *det* “aquilo [gênero neutro]”) e até com dezesseis (línguas bantas). As categorias menos desenvolvidas dentro do sistema flexional parecem ser as de pessoa, comparação e modo verbal (o inglês possui dois modos: indicativo *lives*, imperativo-subjuntivo-infinitivo *live-*, o português possui três: indicativo *vive*, subjuntivo *viva*, imperativo *viva*, o dinamarquês também possui três: indicativo *lever*, imperativo *lev*, subjuntivo-infinitivo *leve-*, alemão e latim, em contraste com o inglês e com o dinamarquês, e assim como o português, distinguem entre o subjuntivo e o infinitivo. Além disso, o grego adiciona a esses dois um optativo separado).

Essa breve investigação é, claramente, tanto incompleta como provisória e tem como intenção apenas sugerir os primeiros resultados que podemos esperar obter a partir de uma tipologia da estrutura linguística. O que eventualmente podemos ganhar dela – e o que naturalmente não pudemos demonstrar nesses exemplos – é o estabelecimento de leis gerais que assinalem quais categorias uma língua *pode* ter, quais categorias *deve* ter, e quais categorias estão ligadas em uma relação implicativa geral.

Ao analisarmos a língua, devemos adicionar ao registro dos elementos também o registro de suas variações. Podemos encontrar dois tipos de variantes: as *variedades*, que são ligadas aos seus arredores por uma relação implicativa relacional, e as *variações*, que variam livremente e possuem apenas relações não-implicativas com o que está ao redor. A significância dessa distinção pode talvez ser observada mais facilmente se

considerarmos a representação dos elementos da expressão na escrita manual. Uma letra pode ter diferentes formas a depender das outras letras com as quais ela está conectada: essas são as variedades. Mas cada variedade pode, por sua vez, tomar diferentes formas a cada vez em que é escrita: essas são as variações. Da mesma maneira, como fora demonstrado pela análise física, existem variedades e variações dos sons falados. Do ponto de vista da estrutura linguística, essa divisão em variantes é universal, isto é, pode ser executada em quaisquer entidades: qualquer entidade possui tantas variedades quanto o número de possibilidades relacionais que tem, e cada uma dessas variedades pode ser posteriormente dividida em um número praticamente infinito de variações. O fato peculiar é que essa divisão em variedades e a divisão em variações se esgotam alternadamente: em uma análise continuada chega-se primeiro a um ponto em que as variedades já não podem mais ser divididas em variedades, mas somente em variações; então chega-se a um ponto em que as variações já não podem ser divididas em variações, mas somente em variedades, e assim por diante. Podemos visualizar isso com mais facilidade ao considerarmos as variantes de entidades do plano do conteúdo. Se tomarmos o conteúdo de um signo como a palavra *homem*, podemos dividi-lo em variedade até, por exemplo, chegarmos à categoria dos homens que podem estar no lugar em que estou agora e, a partir desse ponto, não podemos fazer mais nenhuma divisão em variedades. Porém, podemos dividir essas variedades em variações, o seja, os diferentes homens que podem ficar neste lugar. Esses não podem mais ser divididos em variações, mas em variedades – em cada um desses homens em lugares diferentes. Esses podem ser divididos em variações – em cada um desses homens em cada um desses lugares em diferentes pontos do tempo. Esses podem ser divididos em variedades – como se observados de diferentes pontos de vista. E esses, novamente, podem ser divididos em variações – como se vistos por diferentes pessoas. A uma variação que não pode ser dividida em variações, mas apenas em variedades, daremos o nome de *individual*. A uma variedade que não pode ser dividida em variedades, mas apenas variações, daremos o nome de variedade *localizada*. Como fica evidente pelo exemplo, existem variedades individuais e localizadas de diferentes graus.

Tipologia do uso linguístico

A linguística clássica cultivou particularmente o estudo dos sons (*fonologia*, *fonética*) de pontos de vista tanto fisiológico como físico. Assim, pode-se esperar que uma tipologia *fonética* das línguas poderia ser assentada como parte da tipologia do uso linguístico no plano da expressão da língua. Seu objetivo seria estabelecer que certas comunidades linguísticas usam certos sons enquanto que outras comunidades usam outros certos sons para representar os elementos da expressão de suas línguas. E, é claro, tendo em vista o que dissemos anteriormente, não se preocuparia somente com cada som individualmente, mas com as classes de sons, *categorias* sonoras.

De fato, a fonética clássica caracterizou os sons da fala a partir de muitos pontos de vista diferentes. Dividiu-os em sons pronunciados com a boca mais aberta – o que chamaremos de vocoides – e sons produzidos com a boca mais fechada – os *contoides*. Ambas as categorias podem ser divididas posteriormente de diferentes formas: por exemplo, de acordo com os movimentos dos órgãos de fala usados para produzi-los e de acordo com o lugar em que esses movimentos ocorrem. Dessa maneira, os vocoides foram divididos, por exemplo, em abertos e fechados, anteriores e posteriores, arredondados (produzidos com o arredondamento dos lábios) e não arredondados, nasais e não nasais. Similarmente, os contoides foram divididos também em categorias como labiais, dentais, velares; oclusivas, fricativas, africadas; nasais e não nasais; vozeadas e desvozeadas e assim por diante.

Contra tais divisões, a objeção linguisticamente legítima feita fora a de que essas categorias nunca podem ter quaisquer limites delimitados, seja a partir de um ponto de vista físico (fisiológico) ou linguístico. A partir do ponto de vista físico, observam-se apenas transições contínuas entre essas categorias. Entre os sons abertos e fechados, existe um sem número de sons meio-abertos e meio-fechados. Entre vocoides anteriores e posteriores, pode-se localizar vocoides centras e todos os tipos de outras nuances. Um som pode ser fortemente ou fracamente, completamente ou parcialmente nasalizado, vozeado ou arredondado e assim por diante. Se compararmos os elementos da expressão de diferentes línguas e suas representações fonéticas usuais, veremos que também do ponto de vista linguístico não existem limites universais fixados. Quando aplicamos o teste de comutação aos elementos da expressão em diferentes línguas, torna-se evidente que não existe uma relação fixada entre as fronteiras estabelecidas no domínio fonético pelas categorias da estrutura linguística e as fronteiras que podem ser estabelecidas entre

as categorias *fonéticas* a partir de um ponto de vista físico ou fisiológico. As categorias de vogais e consoantes estabelecidas não coincidem exatamente com a divisão fisiológica em vocoides e contoides. Nas palavras tchecas *vlk* e *krk* (p. 58), *l* e *r* são vogais, mas ao mesmo tempo são contoides (veja também a palavra *little* em inglês). Nas interjeições em muitas línguas, vários contoides podem ser também vogais (observe *pst*, *hm*). Por outro lado, vocoides também podem ser consoantes (como, muito frequentemente, *i* e *u*, que são normalmente representados ortograficamente como *y* e *w* em inglês). Como cada língua estabelece suas próprias fronteiras, as línguas não podem ser congruentes a partir de um ponto de vista fonético, e isso será verdade enquanto continuarmos a análise. Porque o número de elementos da expressão difere de uma língua para outra, claramente não haverá uma relação fixada entre esses elementos a categorias de sons que os representam. Em francês e inglês, o *s*: um vozeado e um desvozeado e cada um deles representa dois elementos da expressão diferentes, como demonstrado pelo teste de comutação (fr. *poison* “veneno” e *poisson* “peixe” diferem na expressão apenas porque um possui um *s* vozeado e o outro um *s* desvozeado; o mesmo ocorre no inglês *zeal* “zelo” e *seal* “foca” e no português *casa* e *caça*), o dinamarquês, porém, não faz tal distinção. Em dinamarquês, o *s* vozeado pode ser substituído em qualquer lugar pelo *s* desvozeado sem nenhum risco de confusão ou mal entendido, enquanto que isso seria impossível em francês, inglês ou português. Em francês e finlandês, *t* e *d* são distinguidos na pronúncia principalmente pelo fato de que *t* é desvozeado e *d* é vozeado. Em dinamarquês, a principal distinção é que *t* é aspirado (seguido por uma espécie de *h*) enquanto que *d* não é. Aqui, novamente, diferentes fronteiras foram assentadas dentro do campo fonético por diferentes línguas, de tal forma que o francófono confundirá um *d* dinamarquês com um *t*, e um dinamarquês, a menos que possua alguma habilidade prévia, será incapaz de ouvir as diferenças entre *d* e *t* em francês ou finlandês. Por outro lado, não surgirá nenhuma confusão ou mal entendido em francês ou finlandês se um *t* aspirado for substituído por um *t* que não o seja e que é comum nessas línguas, ou ainda em dinamarquês se um *d* vozeado for substituído pelo *d* desvozeado usual. O finlandês não faz distinção entre *k* e *g*, ou entre *p* e *b*; seja na escrita ou na pronúncia, normalmente apenas encontra-se *k* e *p*. Como consequência disso, um finlandês será incapaz, sem um treinamento especial, de captar a distinção que outras línguas fazem nesse aspecto; e, por outro lado, será possível substituir o *g* e o *b* dinamarqueses ou o *g* e o *b* franceses pelo *k* e *p* finlandês em qualquer palavra finlandesa sem risco de causar alguma confusão ou mal entendido. Esses são apenas alguns exemplos individuais; a situação é a mesma em todos os lugares – os

sistemas fonéticos das línguas são incongruentes e, por essa razão, é impossível estabelecer um sistema fonético universal que possua qualquer validade linguística.

Pode-se pensar em alcançar melhores resultados ao tomar como ponto de partida algum outro tipo de combinação de estrutura linguística e de uso linguístico diferentes desses que escolhemos. Assumimos que todos os sons concebíveis (fisicamente possíveis) podem representar, ou serem feitos para representar, variantes dos elementos da expressão de uma língua, mesmo que tais sons não sejam costumeiramente ou deliberadamente usados na língua em questão. Dissemos, por exemplo, que *poderíamos* introduzir um *s* vozeado em dinamarquês, ou que em finlandês *poderíamos* introduzir um *t* aspirado no lugar de um que não o seja, ou um *g* e um *b* em vez de um *k* e um *p*. Pode-se objetar que, embora teoricamente pudéssemos fazer isso, de fato, não o fazemos. O *s* dinamarquês, quando pronunciado corretamente, é *sempre* desvozeado, o *t* finlandês é *sempre* não aspirado e o *k* e o *p* finlandeses são sempre *k* e *p*, nunca *g* e *b*.

Essa objeção aparentemente muito plausível, entretanto, envolve, na verdade, uma suposição extraordinariamente precipitada. Como podemos saber essas coisas? É certamente impossível examinar tudo que todos já falaram em dinamarquês ou finlandês, ainda mais o que ninguém ainda não disse, mas irá dizer nessas línguas. Por essa razão, os linguistas em geral sabiamente abstêm-se de afirmar que algo *não pode ser pronunciado* de alguma forma particular em uma determinada língua. Em vez disso, eles se contentam em dizer de que maneiras (implicando: *dentre outras possibilidades*) se *pode* pronunciar. E, mesmo assim, parece haver algo tentador no raciocínio de que dentro de todo o continuum de sons possíveis da fala humana possa haver, em cada uso linguístico, certas terras de ninguém, zonas não ocupadas por qualquer elemento linguístico da expressão. Se assim for, não seria apenas verdade que cada língua estabelece suas fronteiras no domínio fonético, mas também seria verdade que cada língua seleciona seu próprio território em um campo pré-estabelecido. Dessa maneira, poderíamos imaginar um uso linguístico contendo os vocoides anteriores e posteriores, mas sem os centrais, ou vocoides não-arredondados, mas com arredondados. Porém, sempre que empreendemos a investigar um uso linguístico em um pouco mais de detalhe, desapontamo-nos com tais exceções e a proposição teórica inicial não consegue se sustentar de maneira precisa.

Poderíamos tentar um outro caminho: seguindo o foneticista Daniel Jones, poderíamos dividir os sons que surgem como variantes de um único som da fala (um único fonema – isto é, um único representante fonético de um elemento da expressão) em

variantes principais e variante secundárias. Em línguas em que vocoides anteriores e posteriores, mas não centrais, são sons de ocorrência comum, podemos dizer que as diferentes nuances possíveis dos vocoides anteriores e posteriores são variantes principais dos seus respectivos sons da fala (fonemas), enquanto que os vocoides centrais seriam variantes subsidiárias. Analogamente, em línguas em que é comum a ocorrência de vocoides não-arredondados (inglês ou russo, por exemplo), mas os arredondados ocorrem apenas em casos especiais, podemos dizer que os primeiros são as variantes principais, enquanto que os últimos são as variantes subsidiárias dos seus respectivos sons da fala (fonemas). Mas a questão principal aqui é como podemos eliminar a arbitrariedade e a subjetividade dessas decisões: que uma coisa seja de “ocorrência comum” ou “frequente” é meramente uma questão de opinião. Portanto, é improvável que consigamos construir uma tipologia fonética de outra forma senão indicando quais fronteiras a estrutura linguística pode estabelecer no domínio fonético. É no mínimo de se duvidar que pudéssemos declarar certas zonas como terras de ninguém em uma determinada língua e, mesmo assim, chegar a proposições gerais defensáveis sobre uma correlação implicativa entre as categorias fonéticas. Uma tentativa extraordinariamente fascinante nessa direção fora empreendida pelo linguista tcheco-russo Roman Jakobson (na época nos Estados Unidos), que pensa ser possível demonstrar que crianças ao redor do mundo aprendem os sons da fala em uma ordem fixa – primeiro certos sons e depois outros – e que a construção do sistema fonético de uma língua em uma criança será revertido se a faculdade da fala é perdida por causa da afasia como resultado de uma lesão cerebral, o sistema fonético então seria quebrado primeiro a partir dos sons que a criança adquire por último e, em seguida, perdendo os sons que são adquiridos por último. Por fim – e aqui chegamos a um ponto decisivo – afirma-se que a hierarquia fonética resultante se refletirá nos sistemas fonéticos do uso linguístico, de forma que certas línguas possuirão apenas os sons que as crianças aprendem primeiro e que a afasia elimina por último, enquanto outras terão os sons que a criança aprende depois e assim por diante. Com relação aos distúrbios da fala e da língua que as crianças podem ter, dificilmente pode-se fazer objeções teóricas mais desenvolvidas contra essa hipótese (porém, desde que é derivada somente da observação, não pode nem ser confirmada ou invalidada por observações posteriores e é impossível prever qual será seu resultado); até o momento, entretanto, no que concerne ao estudo da estrutura dos sistemas fonéticos, surgem questionamentos teóricos do fato de que essa hipótese pressupõe a existência de uma área pré-forma sem “dono” no uso linguístico. Parte dessa hipótese, com toda certeza, pode ser mantida,

mesmo no que diz respeito ao sistema linguístico de sons, especificamente a parte que envolve uma relação implicativa, não entre sons, mas entre as fronteiras fonéticas linguisticamente determinadas. Por exemplo, embora provavelmente não se possa sustentar que a existência de contoides velares pressupõe a existência de contoides labiais e dentais, pode ser possível manter a existência de, digamos, uma fronteira linguística (isto é, uma fronteira separando os representantes de dois elementos da expressão diferentes) entre as vogais arredondadas anteriores e posteriores (veja o fr. *lu* e *loup*) pressupõe a existência de uma fronteira linguística entre as vogais arredondadas semicerradas anteriores e posteriores (veja o fr. *feu* e *faux*). Mas essa hipótese é, também, inteiramente baseada na observação e deve, portanto, ser confirmada ou invalidada pela observação; e aqui, também, prever o resultado disso é arriscado.

No estudo do significado (*semântica*), esperamos ser capazes de alcançar uma tipologia do uso linguístico para o plano do conteúdo da língua. Isso é, por muitas razões, uma tarefa muito mais difícil do que desenvolver uma tipologia fonética, em parte porque a semântica fora muito menos cultivada e também porque abrange um domínio muito mais amplo. O conteúdo da língua nada mais é que o mundo que nos cerca, os significados específicos e mínimos da palavra, os significados particulares que são individuais (veja a p. 120) são as *coisas* do mundo: a lâmpada que está aqui na minha mesa é um significado particular da palavra *lâmpada*; em mesmo sou um significado particular da palavra *homem*. Mas essas *coisas* organizam-se naturalmente em muitos tipos de *categorias* e outra dificuldade é que mal se sabe que tipo de ciência se preocupa com o estudo do estabelecimento dessas categorias. Com certa justiça, poder-se-ia dizer que todas as ciências reunidas tem esse papel. Todas as ciências, com exceção da linguística, são, na verdade, teorias do conteúdo linguístico estudado independentemente da estrutura linguística, assim como a fonética fisiológica e a física acústica constituem o estudo da expressão linguística independentemente da estrutura linguística. Particularmente, poderíamos sugerir que a psicologia como sendo a ciência, se é que existe uma, que pode fornecer uma tal categorização da coisa experienciada que seria adequada para mapear no sistema de categorias fornecida pela estrutura linguística. É de se esperar que as recentes tentativas de psicólogos em uma descrição “fenomenológica” do mundo externo como imediatamente experienciado possa levar a uma frutífera colaboração com os linguistas. Cabe a nós manter uma posição mais cética em relação a essas tentativas – especialmente aquelas privilegiadas em tempos passados – de fundar uma semântica linguística sobre a base da lógica. Principalmente porque essa perspectiva encerra um círculo vicioso: a

lógica conceptual é baseada na língua (a lógica de Aristóteles, por exemplo, nunca teria a forma que tem se não tivesse sido pensada em grego); a lógica conceptual é sempre uma língua disfarçada e transformá-la ou refiná-la não parece ser produtivo.

As perspectivas para uma tipologia do significado ainda são, portanto, incipientes. E, por isso, devemos antecipadamente ter em mente que ela encontrará as mesmas dificuldades que a tipologia dos sons. Cada língua estabelece suas próprias fronteiras no mundo das coisas e ideias. Um signo em uma língua não corresponde a um signo em outra língua, como todo tradutor sabe. Do ponto de vista do mundo das coisas, assim como do mundo dos sons, não existe congruência entre as línguas. E acreditar que possamos encontrar dentro do mundo das coisas ou das ideias zonas que são variantes subsidiárias ou campos sem dono em uma determinada língua é altamente improvável. Aqui, no que diz respeito à tipologia fonética, será, no melhor dos casos, uma questão de estabelecer relações implicativas, não entre significados, mas entre fronteiras de significado linguisticamente determinadas.

Como as considerações anteriores indicam, uma tipologia do uso linguístico possui dificuldades cruciais para enfrentar. A única tipologia linguística que parece ser exequível em uma base objetiva é, paradoxalmente, aquela que a linguística clássica nunca suspeitou: a tipologia da estrutura linguística.

Mudança linguística

A grande descoberta que fora resultado da pesquisa em linguística genética no século dezenove, e que veio para deixar sua marca na maioria dos linguistas daquele período, foi a de que *as línguas mudam*. Os diferentes membros de uma família linguística devem ser considerados como diferentes transformações de um ponto de partida comum – da língua originária. Na história individual das línguas, também pode-se observar contínuas mudanças e diferentes estágios podem ser distinguidos. Por resultado do interesse desbalanceado que a linguística clássica possuía no uso, ao custo de deixar a estrutura de lado, foi fácil extrapolar a importância dessa descoberta. Pois é óbvio que a estrutura linguística é algo relativamente estável, que pode muito bem no decorrer do tempo, mas mantém-se constante durante longos períodos de tempo, enquanto que o uso é muito mais suscetível a mudanças. Palavras e signos estão incessantemente entrando e caindo em desuso; pronúncias e significados variam de um lugar para outro e mudam de uma década para outra e, é claro, sob uma análise mais profunda, demonstram um constante deslizamento. Para uma ciência que não viu na língua nada a não ser signos, suas pronúncias e os seus significados – e que, além do mais, celebrou seus grandes triunfos sobre as descobertas da mudança linguística – isso estava fadado a levar a quase fatal necessidade de abandonar o conceito de *estado linguístico* (*état de langue*).

Enquanto isso, mais cedo ou mais tarde, essa atitude unilateral estava fadada a encontrar dificuldades, e a própria doutrina das famílias linguísticas e da mudança linguística, quando realizada, levaria a uma compreensão nova e mais profunda do estado linguístico. Fica claro que, por si mesma, a língua originária, conhecida apenas como um conjunto de fórmulas que designam as funções-elemento, é, também, um estado linguístico e, o que é mais importante, um estado linguístico em que a estrutura linguística é o aspecto mais evidente. Mesmo que quiséssemos ir tão longe a ponto de não admitir nenhuma ligação entre as fórmulas indo-europeias e qualquer realidade além das funções-elemento, não se pode negar que quando as fórmulas são reunidas, como fizemos acima nas páginas 47 e 48, elas formam algo que se assemelha, em todos os aspectos, ao sistema da expressão da língua, um sistema de elementos da expressão. Certamente, nossas fórmulas – pseudônimo para os elementos da língua originária – são, cada uma, definidas apenas como representando uma certa função de elementos entre as línguas geneticamente relacionadas. Mas simplesmente não podemos deixar de defini-las também por suas funções *uns com os outros*, ordenando-os em categorias como os

elementos de expressão de qualquer outra língua e dividindo-os em vogais, coeficientes, consoantes, acentos e assim por diante. Para além disso, não é apenas comprovadamente inevitável, mas também se provou altamente frutífero para se estabelecer regras de regência para as possíveis combinações de vogais e coeficientes em ditongos, como também todo um conjunto de regras para a formação de palavras que não são meramente – ou essencialmente – relacionadas aos elementos da expressão do estado linguístico envolvido, mas, antes de tudo, com seus elementos do conteúdo, ou pelo menos com suas unidades do conteúdo e suas possíveis combinações (veja a p. 49). Mesmo que não nos arrisquemos a dar tamanho passo ao atribuir existência real à totalidade de nossas fórmulas, e assim igualá-las a um estado linguístico, não podemos ignorar o fato de que essas fórmulas constituem um *sistema*.

Tal visão sistemática das fórmulas fora primeiramente tomada pelo linguista suíço Ferdinand de Saussure em um trabalho publicado em 1879 e que marca a virada epistemológica na história da linguística, apesar de que o ponto de vista que ele representava era deveras estranho para os seus contemporâneos e por isso teve dificuldades em alcançar uma compreensão geral. De fato, mesmo que a significância prática das descobertas de Saussure seja hoje reconhecida por todos que trabalham com problemas genéticos do indo-europeu, foi somente em tempos mais recentes que nos tornamos completamente preparados para entender o escopo teórico desse trabalho. Sua característica mais distinta é que, por um lado, ela trata as fórmulas como um sistema e tira disso todas as consequências que se seguem desse tratamento, enquanto que, por outro lado, não atribui a elas nenhum outro tipo de realidade e, conseqüentemente, não as considera sons pré-históricos que possuem uma certa pronúncia e que, por uma mudança gradual, tornaram-se sons específicos de cada língua indo-europeia.

Esse ponto de vista levou a avanços práticos na elucidação do próprio indo-europeu. Precisamente porque Saussure trata as fórmulas como um sistema e, ainda, como um sistema independente de qualquer definição fonética concreta – em resumo, como uma estrutura linguística pura – nesse trabalho, ele aplica à língua originária do indo-europeu, a própria fortaleza da teoria da mudança linguística, métodos destinados a estabelecer um padrão de análise de qualquer état de langue, métodos que podem ser tomados como modelos da maneira sob a qual uma estrutura linguística deve ser analisada. Saussure toma o sistema em si e por si mesmo e questiona-se: como posso analisa-lo de forma a obter a explicação mais simples e mais elegante? Em outras

palavras: qual o menor número de fórmulas ou elementos eu preciso para dar conta de todo esse mecanismo?

A partir desse ponto, Saussure foi capaz de fazer coisas com o sistema do indo-europeu que ninguém havia sido capaz antes e, assim, introduziu um novo método, um método estrutural, na linguística genética.

Vejamos um exemplo. Observamos acima (p. 50) que o indo-europeu possui a seguinte alternância vocálica $*e : *o : \emptyset$, a qual aparece em ditongos como $*e\grave{i} : *o\grave{i} : *I$, $*e\grave{u} : o\grave{u} : *u$ etc. Além disso, encontra-se no indo-europeu um tipo diferente de alternância – ou algo que se assemelha a um tipo diferente de alternância – a saber, uma mudança entre uma vogal longa e $*A$. Isso é visto, por exemplo, em

h.a. <i>sthi-táh</i> “de pé”	:	<i>ti-ṣṭhā-mi</i> “eu estou de pé”
lat. <i>sta-tus</i>	:	<i>stā-re</i> “ficar de pé”
raiz do I.E. $*stA$:	$*stā-$

ou em

h.a. <i>di-táh</i> “dado”	:	<i>dā-nam</i> “um presente”, <i>dá-dā-mi</i> ‘I give’
lat. <i>da-tus</i>	:	<i>dō-num</i>
raiz do I.E. $*dA-$:	$*dō-$

(Devemos lembrar que em h.a. $i = \text{Lat. } a$ é em I.E. $*A$; veja a p. 39)

Agora Saussure percebe que se a vogal longa nessas alternâncias for interpretada como uma combinação de uma vogal curta com $*A$, os dois tipos de alternâncias, que antes pareciam totalmente diferentes, tornar-se-iam basicamente as mesmas:

	$*e\grave{i}$:	$*o\grave{i}$:	$*i$
\neq	$*e\grave{u}$:	$*o\grave{u}$:	$*u$
\neq	$*eA$:	$*oA$:	$*A$.

A linha que contém $*A$ torna-se então bastante semelhante com as linhas contendo $*i$, $*u$, ou qualquer outro coeficiente. Todos são casos especiais da linha $e : *o : \emptyset$, e $*A$ deve ser considerado como um coeficiente paralelamente a $*i$, $*u$, $*r$, $*l$, $*n$, e $*m$. Podemos adicionar que fora Saussure quem introduziu o termo “coeficiente” e registramos ainda que no mesmo período Saussure ainda introduziu dois outros termos à linguística, um que

desde então tem encontrado grande aceitação entre os linguistas, a saber, o termo “fonema” (veja a p. 123). Ele o introduziu, por falta de melhor palavra, como uma definição para os elementos da expressão de uma língua para assim evitar confusão com os “sons” do uso linguístico – ou seja, para designar puramente “as entidades algébricas” de sua teoria. Por ironia do destino, a teoria de Saussure foi lida de maneira tão rasa e por isso mal entendida por seus contemporâneos – e por muitos que vieram depois – que o termo é geralmente usado como um sinônimo de “som linguístico” – precisamente o que Saussure pretendia evitar.

Porventura, essa descoberta de Saussure não soe muito impressionante. Se assim o for, a razão é que, na discussão feita acima, o material foi organizado para o leitor e, assim, a solução tornou-se bastante óbvia. Entretanto, se considerarmos as noções comuns da época, o efeito gerado provavelmente foi um tanto surpreendente, mesmo porque, como vimos (p. 49), as vogais curtas **e* e **o* podem também alternar com as vogais longas **ē* e **ō* (veja, por exemplo, o nominativo grego *rhētōr-a* “orador”, com **ō* e o acusativo *rhētōr-a* com **o*) de forma que a vogal longa que está em alternância com **A* possa estar funcionalmente em uma linha com as vogais longas alternando com zero e não, como se viu, com as vogais curtas correspondentes.

Essa reinterpretação das vogais longas alternando com **A* como uma vogal curta + **A* foi possível, é claro, apenas porque a combinação de uma vogal curta + **A* não aparece de outra forma no indo-europeu; no entanto, significou uma ruptura brusca com o método anterior de reconstrução porque uma fórmula como a **o_A* de Saussure é fundada, não nas funções-elemento entre as línguas indo-europeias, mas nas funções internas à língua originária. Se as evidências estivessem limitadas as funções-elemento entre as línguas indo-europeias conhecidas no tempo de Saussure, não haveria motivo para se distinguir o *ō* em *dō-num* e o *ō* em *rhētōr*. Se o *ō* em *dō-num*, mas não o *ō* em *rhētōr*, pode ser interpretado como **ao*, não é por causa de uma função entre diferentes línguas, mas por causa de uma função entre constituintes de um único estado linguístico. O que aconteceu aqui é que uma entidade algébrica fora equalizada com o produto de duas outras; essa operação relembra a do químico que analisa água como um produto do hidrogênio e do oxigênio. Esse tipo de operação é exigida pela análise de qualquer estado linguístico se se quiser alcançar a descrição mais simples possível. Paradoxalmente, essa operação fora aplicada primeiramente na história da ciência linguística à língua originária do indo-europeu.

Em um artigo publicado ao mesmo tempo que o de Saussure, e independentemente deste, o linguista dinamarquês Herman Møller desenvolveu ideias semelhantes e descobriu também uma nova e importante simplificação: a oposição entre as três vogais longas $*\bar{e}$, $*\bar{o}$, e \bar{a} pode ser interpretada, não como Saussure as interpretou, ao usar um único coeficiente $*_A$, mas usando três. Assim, em vez de $*\bar{e} : *_A : *\bar{o} : *_A$, $*\bar{a} : *_A$, é possível escrever $*_{A1} : *_A1$, $*_{A2} : *_A2$, $*_{A3} : *_A3$. Posteriormente, Møller identificou que certas vogais curtas – $*e$, $*o$, $*a$ – poderiam ser explicadas de maneira semelhante como produtos dos mesmos coeficientes e da mesma vogal: $*e = *_{A2}a$, $*o = *_{A2}a$, $*a = *_{A3}a$.

Para entender o que é essencial e importante dessas reduções de um ponto de vista metodológico, devemos perceber que elas constituem um tipo de resolução das entidades indo-europeias em produtos algébricos ou químicos. Isso não advém diretamente da comparação de diversas línguas indo-europeias, mas de um tratamento posterior dos resultados de tais comparações – a partir de uma análise desses resultados. *Posteriormente*, muito depois que análise fora feita, descobriu-se que existe uma língua indo-europeia que distingue entre $*\bar{o}$ em alternância com $*o$ e $*\bar{o}$ em alternância com $*_A$, a saber o hitita, na qual o linguista polonês Kuryłowicz foi capaz de demonstrar que h , por vezes, corresponde ao $*_A$ do I.E. Além disso, Møller foi capaz de confirmar sua teoria ao apoiar-se no hamito-semítico: que essa língua possua consoantes especiais que correspondem aos diferentes coeficientes do indo-europeu, é, de fato, a pedra basilar na demonstração da relação genética entre o indo-europeu e o hamito-semítico. Essas confirmações, obtidas por meio da consideração de funções-elemento previamente desconhecidas entre línguas geneticamente relacionadas, são, certamente de grande interesse, especialmente por mostrarem que a análise interna da estrutura linguística como essa da língua originária do indo-europeu traz consigo fortes implicações realistas. Pode-se pensar que esse tipo de análise nos desviaria a um mundo de abstrações, mas vê-se o contrário, na verdade, ela nos coloca no foco das funções-elemento ainda não descobertas. Através da análise do estado linguístico, nós verdadeiramente alcançamos um entendimento mais profundo da estrutura linguística. Por outro lado, essas confirmações advindas da análise do hitita e do hamito-semítico são ainda meras confirmações, e a análise interna do sistema de elementos da língua originária poderia ser alcançada sem referência a elas.

Na esteira das descobertas de Saussure e Møller sobre a estrutura da língua originária, várias outras se seguiram, algumas feitas pelos mesmos dois estudiosos e algumas por outros teóricos. O princípio analítico envolvido, de acordo com o qual uma

tentativa é feita com o intuito de reduzir ao mínimo o número de fórmulas-elemento necessárias em cada estágio do desenvolvimento de uma língua – um método que vem sendo seguido no estudo de todos os estados linguísticos desde sua primeira aplicação na língua originária do indo-europeu – contribuiu para a introdução de uma técnica completamente nova dentro da linguística genética e comparativa indo-europeia. No capítulo sobre relação genética, ativemo-nos ao que fora feito na linguística clássica, que foi marcada pelo registro mecânico das funções-elemento entre as línguas, mas o sistema de fórmulas com o qual terminamos (p. 47) tem sido significativamente reduzido por esse novo princípio de análise. Entretanto, iríamos nos afastar bastante se entrássemos em detalhes das extensas reduções que resultam desse procedimento, por mais interessantes que eles sejam. É suficiente mencionar que tem sido possível explicar a combinação de consoante + *h* como uma combinação de consoante + coeficiente $*_A$, e explicar as vogais longas que alternam com $*e$ e $*o$, como também com o zero com o qual elas alternam, além de oposições acentuais (circunflexo e agudo; veja a p. 43) como variedades condicionadas pela localização do acento da palavra e pelo número de sílabas latentes. Reduções analíticas desse tipo levaram a um sistema extraordinariamente simples das fórmulas indo-europeias.

Ainda outras razões destacam a importância do estado linguístico e permitem que ele reivindique seus direitos em face de uma consideração muito exclusiva das transformações da linguagem. A melhor maneira de demonstrar isso é, talvez, toma-la como uma mudança linguística em si mesma. Caso – pelo bem da simplicidade, usaremos um exemplo artificial – encontremos uma língua na qual *p* tenha se tornado *f* no começo de palavras, mas tenha se mantido como *p* em outras posições, deve ter havido um período na história dessa língua em que todo *p* em início de palavra tenha mecanicamente mudado para *f*. Isso seria uma *lei sonora* no sentido verdadeiro da expressão, que seguiria a mesma linha de qualquer outra lei presente na sociedade como, por exemplo, uma lei jurídica: a partir do momento em que a lei entra em efeito até o momento em que cessa seu exercício, estava em vigor e implica a conversão mecânica do *p* inicial em *f* em qualquer palavra que entre nos domínios da sociedade, seja essa palavra herdada, emprestada ou formada recentemente. É precisamente por isso que as palavras emprestadas podem diferir em sua forma externa de uma palavra herdada: elas foram tomadas de empréstimo depois que algumas leis deixaram de existir. Mas tal lei sonora é realmente uma lei de mudança ou uma lei de estado? Se uma sociedade possui uma lei que decreta que sob certas circunstâncias todo homem jovem de uma certa idade deve se tornar um soldado, no

futuro poderá se dizer que todos os homens nessas condições se tornaram soldados? Com certeza, mas a mudança referida nesses termos possui sua causa em um estado de coisas. Contanto que o mesmo estado continue – enquanto a lei vigorar – seria incorreto falar de qualquer mudança. Se pode falar, no entanto, sobre uma conversão mecânica de um elemento em outro sob certas condições. Enquanto a lei estiver em vigor, todo *p* em início de palavra é mecanicamente convertido em *f*. Esse é o mesmo tipo de lei de conversão que a lei que rege um peão no xadrez: quando atinge o outro lado do tabuleiro, pode ter seu valor convertido para aquele de uma rainha. A “mudança” ocorre, ou pode ocorrer, sempre que jogamos xadrez, e irá, ou poderá, acontecer caso joguemos xadrez amanhã ou daqui a um ano. Não é uma mudança: é um estado.

O leitor que não é um linguista profissional provavelmente ficará surpreso que o autor possa discorrer tão acaloradamente sobre um assunto tão simples. A razão para isso é que apenas agora os linguistas tem tido a oportunidade de trabalhar livres de uma perspectiva monocêntrica da mudança. De fato, muito do que seria considerado como uma visão monocêntrica da mudança linguística deve ser entendido, na verdade, como um fenômeno estático. Demonstramos acima (p. 37) que função-elemento especial pode ser estabelecida para I.E. **m* no final de palavra: nessa posição em particular, o grego, por exemplo, possui um *-n* em vez de um *-m*. Mas isso, na verdade, não deve ser percebido como uma função-elemento distinta. Se primeiro prestarmos conta, como devemos fazer, das leis estáticas que prevalecem em cada língua, então essa função-elemento especial é supérflua. O fato de que em grego “*-m* se torna *-n*” não se caracteriza como uma mudança histórica pela qual o grego passou, mas sim uma regra do “jogo” que vale no sistema grego: na estrutura da língua grega, *m* é um elemento da expressão que é parcialmente definido por sua incapacidade de estar em final de palavra e, portanto, sempre que cair nessa posição proibida, deve ser substituído por *-n*. Isso não é mais surpreendente do que o fato de que temos que converter nosso dinheiro sempre que cruzamos a fronteira entre países. Devemos fazê-lo porque cada moeda pode ser usada somente em seu próprio país, e isso não se caracteriza como uma lei de mudança. Assim, a mudança linguística pressupõe o estado linguístico e, portanto, a linguística genética pressupõe a teoria da estrutura linguística, ou gramática, o que, de fato, se refere a linguística tipológica. Apenas depois de termos completamente descrito a estrutura linguística em todos os seus estados linguísticos individuais, e identificado tudo o que pode ser colocado no espaço em que ocupam, é que poderemos justificar uma comparação de um ponto de vista

genético. Dessa forma, a linguística genética será simplificada e esclarecida a um nível nunca antes imaginado.

Mas mesmo que se dê um grau de importância ainda maior ao estado linguístico e se diminua essa carga da mudança linguística, o fato que permanece é o de que a língua, de fato, muda. O linguista deve, então, buscar uma explicação casual para essa mudança. Essa questão disputada e nem um pouco clara é consideravelmente simplificada se nos apegarmos a distinção entre a estrutura linguística e o uso linguístico. Existem muitas razões para as mudanças no uso linguístico, para mudanças na pronúncia (leis sonoras em um sentido mais restrito), mudanças no significado, mudança nos signos (palavras). O ser humano é uma criatura enigmática e capciosa e, aqui, é o seu ateliê. A única declaração provisória que pode ser feita com alguma certeza é que as mudanças sonoras que uma língua passa no decorrer do tempo podem, recorrentemente, ser resumidas em certas *tendências* predominantes, que uma determinada população pode manter por séculos com uma tenacidade notável. Maurice Grammont demonstrou de maneira muito clara como essas tendências, em relação a mudança de pronúncia, em certas direções tem estado ativas nas línguas indo-europeias orientais, desde tempos passados até a modernidade, constantemente tomando para si um novo material suscetível a mudanças. O que está por trás dessas tendências é outra questão. Elas podem ser hábitos de natureza psicológica. Encontramos tais tendências na vida cotidiana sempre que ouvimos um estrangeiro falar nossa língua nativa. O “falar com um sotaque” significa exatamente que por causa das tendências de pronúncia, seus hábitos psicológicos, ele desloca o sistema sonora da língua em certas direções. Essas tendências são adquiridas, naturalmente, com o tempo e não são inatas. Porém, as dificuldades ligadas ao aprendizado da pronúncia de uma língua estrangeira da maneira como um nativo falaria, demonstram que esses hábitos adquiridos são praticamente inerradicáveis e pode-se, portanto, entender como eles podem ser insistentemente mantidos de geração em geração por longos períodos de tempo.

As razões para as mudanças na estrutura linguística, por outro lado, não podem ter como base as tendências dos falantes. A estrutura linguística, afinal, é definida como algo independente delas. Sabemos que a estrutura linguística segue suas próprias regras e que uma mudança em sua estrutura não é decorrente de tendências estabelecidas pelos falantes, mas sim de *disposições* no sistema que sofre a mudança. Uma determinada estrutura pode ser vista como predisposta a mover-se em certas direções e não em outras. E aqui temos uma tipologia linguística, com o que ela tem a dizer sobre categorias que se

procuram e se favorecem e categorias que se evitam ou evitam umas às outras, as quais devem identificar as causas da mudança linguística.

Diferentes graus linguísticos

Toda ciência tem como objetivo prover um procedimento por meio do qual se possa descrever objetos de uma determinada natureza. Isso sempre é feito ao introduzir-se uma *língua* por meio da qual os objetos possam ser descritos: um conjunto de termos é introduzido – uma terminologia acompanhada por definições – e, então, a descrição é feita utilizando-se desses termos para formar sentenças de acordo com o objeto em questão.

A linguística tem por objetivo prover um procedimento por meio do qual as línguas sejam descritas. Isso é feito ao introduzir-se uma língua por meio da qual as línguas possam ser descritas. Tal língua feita para descrever outras línguas é chamada de *metalinguagem*, enquanto que a língua descrita é chamada de *objeto linguístico*. Graças ao universalismo das línguas cotidianas, uma língua desse feitio pode ser usada como metalinguagem para descrever a si mesma como um objeto linguístico: é possível, por exemplo, escrever uma gramática do dinamarquês em dinamarquês. Entretanto, em geral, será necessário mudar um pouco o *uso* da língua cotidiana ao introduzir-se novos signos, os assim chamados termos técnicos. Também é possível substituir a língua cotidiana, totalmente ou em partes, por uma língua simbólica especialmente feita para servir como metalinguagem na descrição de outras línguas. Isso também é feito na linguística, mas não utilizamos esse tipo de metalinguagem neste livro.

Assim, vemos podem existir línguas de diferentes graus, línguas de primeiro grau, de segundo grau, ou até metalinguagens. Teoricamente, é claro, poderíamos continuar nessa progressão: uma língua que descreve uma metalinguagem será uma língua de terceiro grau, ou uma metalinguagem de segundo grau (também chamada de meta-metalinguagem). E veremos que essa possibilidade teórica é realmente verdadeira.

A gramática, ou descrição, do estado individual de uma língua é, portanto, uma metalinguagem de primeiro grau. Ao usarmos os termos introduzidos em vários pontos dos capítulos anteriores, podemos agora resumir brevemente o procedimento gramático ao dizermos que ele principia em uma cadeia ou um texto como objeto de análise, deduz disso uma pressuposição unilateral entre a cadeia (texto) e os paradigmas (a cadeia pressupõe os paradigmas) e, por fim, com base nessa relação, estabelece os paradigmas.

Como todas as ciências (mas ao contrário das línguas cotidianas), a gramática deve definir seus próprios signos o máximo possível. Mas qualquer ciência (e isso inclui a gramática), não importando quantos signos ela defina, será forçada a terminar o seu

conjunto de definições em algum lugar, resultando que os signos que entram nessas definições básicas pressupostas não serão definidos. Assim, em qualquer ciência, restarão alguns pressupostos básicos indefiníveis, ou conceitos básicos – termos que não são definidos dentro da língua científica, mas que somente podem ser definidos através da introdução de outra língua, que será uma metalinguagem em relação a essa língua. Assim, também na gramática restarão certos conceitos básicos, certos termos indefinidos que só poderão ser definidos através da introdução de uma metalinguagem de segundo grau. Desde que todos os termos definidos de qualquer ciência devem necessariamente ser definidos por meio de outros termos tirados da mesma língua, os termos definidos então serão fixados por suas relações mútuas e nunca podem ser fixados (ou mesmo parcialmente fixados) por suas relações com quaisquer outros objetos que não termos dentro da mesma língua. Ao utilizar-se de glosas de uma língua cotidiana, a gramática terá a vantagem sobre as outras ciências de, quando aplicada à língua cotidiana que forma sua base, ela será capaz de definir todos os termos que foram tirados dessa língua cotidiana. Porém, como vimos, ela somente pode defini-las por suas relações com outros termos que pertencem a mesma língua, não (seja em parte ou no todo) por suas relações com objetos que não sejam eles mesmos termos internos a essa mesma língua. Desse modo, os termos da gramática não podem ser fixados por quaisquer coisas ou objetos que pertençam a uma realidade exterior à própria gramática a qual eles se referem ou do qual são feitos (por exemplo, as vibrações eletromagnéticas causadas por marcas de tinta no papel, ou as ondas sonoras causadas pelos movimentos na garganta e na boca). Esses conceitos, por outro lado, podem ser alcançados por uma meta-gramática, na qual analisamos os termos da gramática e identificamos que as coisas as quais eles se referem são fenômenos físicos ou outros fenômenos que perfazem a expressão e o conteúdo do objeto linguístico (a língua cotidiana) e que as coisas com as quais são feitos são as mesmas coisas que formam os termos de outras línguas (a língua cotidiana). Em outras palavras, a gramática, como qualquer língua, deve ser dividida em um plano do conteúdo e um plano da expressão. O que forma o conteúdo de uma gramática, o plano do conteúdo, é a expressão e o conteúdo de uma língua cotidiana. Para tratar dentro de seu plano do conteúdo com os elementos da expressão e os elementos do plano do conteúdo da língua cotidiana, a gramática é obrigada a introduzir termos para eles. Pode ocorrer, por exemplo, de um elemento da expressão ser chamado de *p*, outro de *m* e assim por diante; e pode-se chamar um elemento do conteúdo de “ele”, outro de “ela”, outro, talvez, de “bovino” e assim por diante. Mas todos esses termos introduzidos pela gramática para os

elementos da língua cotidiana continuarão como conceitos básicos indefiníveis, que serão definidos somente em uma meta-gramática por meio de uma análise do plano do conteúdo dessa gramática. Desse modo, a fonética, ciência dos sons da língua, e a semântica, ciência dos significados da língua, pertencem a uma meta-gramática. Entre uma meta-gramática e uma gramática (bem como entre qualquer metalinguagem e seu objeto linguístico) existe uma relação unilateral de pressuposição: a metalinguagem pressupõe o objeto linguístico e a meta-gramática pressupõe, então, a gramática. Em outros termos: sons e significados pressupõe os elementos da expressão e do conteúdo; a representação pressupõe o que é representado.

Mas existe outra forma pela qual podemos encontrar diferentes graus linguísticos. Se realizarmos uma análise gramatical de um determinado texto, desde que o texto seja longo o suficiente (a literatura dinamarquesa em conjunto com uma coleção de gravações do dinamarquês falado), veremos que dentro de todo o conjunto textual podem surgir – ou por extensão do texto podem ser produzidos – fragmentos (componentes textuais grandes, ou palavras individuais, ou elementos individuais, ou variantes individuais) que podem ser traduzidos dentro do texto em outros fragmentos. Um fragmento, talvez, esteja no dialeto jutish e outro no dialeto zeelandico; verificamos então ser possível traduzir o fragmento jutish em zeelandico, e o zeelandico em jutish. Ou um fragmento está escrito enquanto que o outro é falado e ambos podem ser mutuamente traduzidos. Ou um fragmento é a fala cotidiana, enquanto que a outra é a fala no estilo de uma palestra; um fragmento é prosa, outro, poesia – em todos os casos, traduções mútuas são possíveis. Se tivéssemos um texto incluindo diferentes línguas de diferentes países, como o inglês e o dinamarquês, teríamos a mesma situação. No momento em que tivermos um texto registrado por duas pessoas diferente, teremos, novamente, a mesma possibilidade de tradução. Toda fisionomia linguística (“voz”, “caligrafia”) possui sua própria marca no uso linguístico e um texto pode ser traduzido de uma fisionomia para outra. Se colocássemos um certo número de estudantes para ler em voz alta, ou relatar, ou escrever a mesma história, eu obtenho um número de texto traduzidos em diferentes fisionomias.

A cada vez que encontrarmos tal tradutibilidade entre dois fragmentos de um texto, devemos considera-los na análise. Sempre que encontrarmos diferentes línguas, dialetos locais, jargões, expressões idiomáticas, estilos, fisionomias, devemos analisar separadamente esse texto que cada um traz consigo, porque ele pode revelar diferenças na estrutura linguística. Assim, na gramática, elaboramos para nossa análise um sistema de rotulação, de forma que um grupo de texto seja marcado como “inglês”, “dinamarquês”

ou “português”, ou como “prosa” e outro como “poesia”, ou um como “John Smith” e outro como “James Brown” e assim por diante – tais divisões, obviamente, cruzam-se de diversas formas. Mas os rótulos em si mesmos – “inglês”, “dinamarquês”, “português”, “prosa”, “poesia” etc. – não podem mais ser tratados dentro da linguagem de nossa gramática.

O que é curioso, de fato, é que essas denominações representam conceitos que são, em geral, sagrados e que envolvem o emprego de um determinado uso linguístico ou de uma determinada estrutura linguística como expressão de um conteúdo formado por certos elementos que estão fora da língua. Assim, o português brasileiro é a expressão do conteúdo que representa a nação brasileira, a família, a moradia; e, da mesma forma, diferentes estilos são expressões, ou símbolos, para conteúdos formados por certos elementos que estão fora dos estilos. Então, novamente, estamos lidando com uma expressão e com um conteúdo – com a língua – mas dessa vez com uma língua cujo plano da expressão é, ele mesmo, uma língua, com seu próprio conteúdo e expressão. Desse modo, devemos adicionar à nossa gramática uma nova estrutura gramatical, que trata de todos esses “rótulos” (ou conotadores, como os chamaremos) como conteúdos para uma determinada língua que é a expressão (chamaremos essa função conteúdo-expressão de conotação). Então, novamente, teremos que lidar com gramáticas de diferentes graus, mas em um sentido diferente daquele visto anteriormente. Pode ser apropriado continuar por esse caminho: em uma gramática de terceiro grau, poderemos demonstrar que elementos geográficos, históricos, sociais e psicológicos compõem o conteúdo do nome da língua – “português brasileiro”, por exemplo – considerados como expressões para tal conteúdo. Tudo, desde características nacionais até as mais pessoais, encontra um lugar natural dentro da esfera linguística.

As tipologias genética e linguística são, é claro, línguas assim como a gramática; elas são meramente extensões de seus domínios. Cada uma, como vimos, tratam de certas funções entre estados da língua. Para usar os termos que introduzimos acima, podemos resumir ao dizer que a linguística genética estabelece as línguas originárias, enquanto que a linguística tipológica estabelece as tipologias por trás das línguas individuais. Existe uma correlação de pressuposição unilateral entre as línguas individuais e a língua originária: as línguas individuais pressupõem a língua originária, mas não o contrário. E a função entre as línguas individuais e a língua originária é uma correlação, não uma relação: elas não estão presentes conjuntamente, mas em alternância. Entre as línguas individuais e a tipologia linguística existe uma relação de pressuposição unilateral: as

línguas individuais pressupõem a tipologia linguística, mas não o contrário. E a função entre uma língua individual e uma tipologia linguística é uma relação, não uma correlação: a língua individual e a tipologia linguística estão presentes conjuntamente; se nos é dada uma linguagem individual, então nela, *eo ipso*, nos é dada sua tipologia. A essa altura, podemos finalmente alcançar uma *definição formal* dos dois tipos de relação linguística: a relação genética se assenta sobre uma continuação, ou seja, uma correlação de pressuposição unilateral entre línguas, enquanto que a relação tipológica se assenta sobre uma realização tipológica, ou seja, uma relação de pressuposição unilateral entre línguas.

Índice remissivo

- * A, 38, 43, 96, 126-28
 Abreviação dos signos, 71-74
 Acústica, 122
 Afegão, 86
 Africâner, 83
 Ainu, 91
 Acádios, 86
 Albanês, 77, 85
 Aleúte, 91
 Altaico, 90, 101
 Alternância, 48, 126-28
 Análise, 29, 63, 104, 108, 117, 124, 127
 Anglo-frísio, 82
 Afasia, 121
 Árabe, 86-87
 Aramaico, 89
 Austríaco, 91
 Austro-asiático, 92
 Avéstico, 92
 Antigo cananeu, 93
 Análise universal, 120
 Alternância vocálica, 50, 129
 Alto-alemão médio: citado em, 68
 Alto-alemão, 45, 68, 87; mudança sonora, 46; citado em, 46
 Armênio, 34, 36, 41 citado, 33
 Acento, 35, 36, 42, 67, 95, 105, 116, 128
 Alto-alemão antigo, 33; citado em, 40, 45, 65
 Alemão, 33, 44, 46, 47, 76, 78, 76, 78, 83, 94, 119; citado em, 45, 56, 71, 72, 73, 75,
- 59
- “Bokmål,” Norueguês, 84
 Baixo-alemão, 46, 78, 80, 85
 Báltico, 86, 87
 Banta, 90, 93, 94, 119
 Basco, 90
 Bielorrusso, 87
 Bretão, 85
 Britânico insular, 85
 Britânico, 85
 Búlgaro, 33
 Bushman, 94
 Finlandês báltico, 91
- Cadeia, 53, 106, 136
 Campo, 66
 Canaanita, 88
 Carélio, 91
 Caso, 105, 119, 121-22
 Catalão, 86
 Categoria, 54, 56-58, 73, 74, 105-106, 107-108, 110, 117, 121, 122, 125-26
 Celta, 34, 80
 Chaldean Pehlevi, 87
 Chinês, 87, 90, 92, 102, 104, 118

- Chuvaches, 91
 Ciência, 27, 97, 106, 125, 136-38
 Classe, 31
 Código, 60, 63
 Coeficiente, 48, 131
 Coisas, 125
 Comparação, 119
 Comutação, 109, 118, 119
 Conotador, 139
 Consoante, 46, 54, 106, 107, 117-19, 122
 Constante, 59
 Contaminação, 72
 Conteúdo, 52, 73, 99, 106, 108, 110, 111, 125, 139
 Contoide, 117, 123, 125
 Contraexemplos, 50-51, 63, 64, 72, 73, 81, 83
 Cótico, 89, 93
 Coreano, 93
 Córico, 85
 Correlação, 53; implicativa recíproca, 114; implicativa relacional, 114, 117, 139;
 não-implicativa, 114
 Cuneiforme, 34, 89
 Curdo, 89
 Cuxítico, 89
- Dálmata, 87
 Definição do conteúdo, 112
 Definição expressiva, 112
 Definição operacional, 50
 Definição, 50, 112, 138
 Dependência. *Veja* Função
 Descrição, 31, 126, 137
 Dinamarquês, 47, 76, 81, 82, 115, 119, 121; citado, 40, 45, 46, 65, 67, 72, 73, 75,
 76, 77, 99
 Direção, 114
 Disposições, 136
 Dissimilação, 66, 82
 Ditongo, 47
 Divisão. *Veja* Análise
 Dravídico, 93
- EE. *Veja* Elemento da Expressão
 EE-função. *Veja* Função-elemento
 Egípcio, 89, 93
 Egípcio antigo, 93
 Elemento básico, 114, 118
 Elemento caracterizante, 114
 Elemento central, 116
 Elemento coberto, 65
 Elemento da expressão, 35, 37-39, 51, 52, 53, 57, 97, 109-10
 Elemento derivacional, 117
 Elemento descoberto, 65
 Elemento do conteúdo, 52, 109, 113 continuação, 50, 139
 Elemento extenso, 116
 Elemento flexional, 114
 Elemento intenso, 1160

- Elemento periférico, 109-11
 Elemento radical, 110
 Elemento, 97-98, 106. *Veja também* básico; central; caracterizante; do conteúdo; coberto; derivacional; da expressão; extenso; flexional; intenso; periférico; radical; radico; descoberto.
 Empréstimo, 59-65. *Veja também* palavras emprestadas
 Escandinavo, 11, 22, 24, 25, 60, 65, 68-69
 Esloveno, 71
 Esloveno, 71
 Espanhol, 22, 70; citado em, 22, 46, 49, 53, 58
 Espécies de elementos, 109
 Esquimó, 81
 Estado linguístico. *Veja État de langue*
 Estônio, 67, 77
 Estrutura de palavras: tipologia, 91-93
 Estrutura, linguística, 7, 31, 32-44, 53, 57, 58, 60-61, 91, 95, 96, 122, 130-31; tipos, 97-114
État de langue, 8, 9, 97, 122-24, 127-30, 133, 136
 Etimologia popular, 54, 60
 Etimologia, 90; popular, 54, 60
 Etiópia, 76
 Etrusco, 81
 Evenki, 78
 Expressão, 4, 95, 99, 101, 102, 108-12, 134, 136
- Família linguística, 9, 21-22, 70-81, 122, Sentimento linguístico. *Veja Sprachgefühl*
 Fenício, 73
 Fenomenologia, 121
 Feroês, 70
 Finlandês, 39, 65, 77, 110, 117-18; citado, 63
 Fino-mordoviano, 77
 Fino-permiano, 77
 Fino-úgrico, 76-77
 Física, 4-5, 47, 113, 115, 120, 134
 Fisiologia, 4-5, 47, 48, 115, 120
 Fisionomia linguística, 135
 Flamengo, 71, 72
 Fonema, 118-19, 125
 Fonética, 4, 115-20, 134
 Formação analógica, 51-56
 Formação de palavras, 28-30
 Formação do signo, 34, 42, 45-67
 Formação proporcional, 55-56
 Fórmula, funcional, 15-30, 81
 Francês antigo, 60; citado em, 56, 66
 Francês, 22, 39, 46, 61, 64, 65, 94, 109, 110, 112, 116-17; citado em, 49, 54, 56, 57, 58, 59, 116, 120
 Frígio, 71
 Frísio, 69
 Função categorial, 95
 Função, 8-9, 13, 15, 32, 33. *Veja também* Função categorial, Função-elemento

- Função-elemento, 13, 15-26 45, 46, 48, 61, 64, 80, 84, 86, 87, 95,122-23
- Gaélico escocês, 69
 Gaélico, 69
 Galego, 70; citado, 49
 Galês antigo: citado em, 18
 Galês, 11, 71; citado em, 14, 18,64
 Galo-romance, 70
 Gaulês, 69
 Generalização, 62
 Gênero, 95, 111-13
 Georgiano, 38-39; citado em, 82
 Germânico comum, 22
 Germânico, 22, 24, 46, 53, 54, 64, 65, 68-69, 72; mudança sonora, 19-20, 24, 87
 Gótico, 11, 22, 68; citado em, 11, 12, 13, 14, 18, 19, 20, 23, 24, 45, 51-52, 53, 54,
 56, 88
- Gramática, 8, 130, 133
 Grammont, Maurice (1886-1946), 51, 130
 Grande russo, 72
 Graus linguísticos, 132-37
 Grego, 11,63,64,67,71, 112, 121, 129- 30; citado, 11, 12, 13, 14, 15, 17, 18, 19, 20,
 21, 28, 29, 45, 50, 51-52, 82, 88, 126
 Groenlandês, 92-93, 94; citado em, 93
- Hamítico, 73
 Hamito-semítico, 73-76, 79, 80, 127
 Haplologia, 50
 Harmonia vocálica, 110
 Hebraico, 73-76
 Helênico, 71
 Hieroglifos, 73
 Hipóteses, 85, 87
 Hispano-romance, 70
 História, 4-5
 Hitita, 11, 73,127; citado em, 14,18,51-52
 Homburger, Lílias, 79
 Hotentote, 81
 Humboldt, Wilhelm von (1767-1835), 93
 Húngaro, 76-77, 112; citado em, 63
- Iacuto, 77
 Ídiche, 76
 Indefiníveis, 133
 Índico antigo, 11, 72; citado em, 11, 12, 13, 14, 15, 17, 18, 19, 20, 45, 48, 50, 51, 52,
 53, 88, 124-25
 Índico médio, 72
 Índico, 11, 72
 Individual, 114
 Indo-europeu, 10, 68-73, 79, 80, 92, 93
 Inflexão, 92, 93
 Inglês antigo, 11; citado em, 23, 24
 Inglês médio: citado em, 18
 Inglês, 11, 22, 59, 65, 69, 71, 76 86 94, 112, 116-17, 119; citado em, 23, 24, 32, 46,
 50, 53, 54, 55, 56, 58, 59 60, 61, 64, 66
 Inscrições de Orkhon, 77

- Inscrições rúnicas, 68
 Iraniano, 72
 Irlandês médio: citado em, 18
 Irlandês, 11, 69; citado em, 11, 12, 13, 14, 17, 18, 19, 20, 45, 51, 64
 Islandês, 64, 68; citado em, 24, 63
 Italiano, 22, 70, 71; citado em, 22, 46, 49, 50, 53, 56
 Itálico, 22, 69-71
 Ítalo-Sardenho, 70
 Iugoslavo, 71
- Jakobson, Roman, 119
 Japonês, 80
 Jogos, 42-43, 99, 101
 Jones, Daniel, 118
 Junção, 107
- Khoisan, 81
 Kierkegaard, Søren (1813-55), 104
 Komi, 77
 Kuhn, Ernst (1846-1920), 79
 Kurylowicz, Jerzy, 127
- “Landsmil,” Norueguês, 68-69
 122, 124, 128, 130, 136
 Lapônico, 77
 Latência, 98, 128
 Latim vulgar, 22, 46
 Latim, 11, 22, 47, 59, 60, 63, 64, 65, 70, 84, 86, 112; citado em, 11, 12, 13, 14, 17,
 18, 19, 20, 32, 46, 48, 49, 50, 51-52, 53, 54, 56, 57, 63, 64, 66, 88, 92, 124-26
 Lei sonora, 87, 128-30
 Lei, 128-29
 Lepôntico, 70
 Letônio, 71
 Letras, 13, 40-41, 43, 86-87, 113
 Líbico-berbere, 73
 Lício, 73
 Língua cotidiana, 101-2, 104-5, 106, 109, 132-33
 Língua de uma criança, 119
 Língua originária, 82-90, 122-23, 136
 Língua, 36-37, 101, 104, 132. *Veja também* Língua cotidiana
 Línguas americanas, 21, 39, 81, 93 Amárico, 76
 Línguas australianas, 80
 Línguas caucasianas, 80
 Línguas originárias do indo-europeu, 26- 27, 85-89, 123-24, 127-28
 Línguas paleoasiáticas, 80
 Línguas românicas, 22
 Línguas sudanesas, 79-80, 81
 Linguística clássica, 3, 6-7, 26, 85, 87, 91, 94, 96, 106, 115, 121, 122
 Linguística Genética, 6-7, 85, 89-90, 96,
 Linguística, 3-7, 96. *Veja também* Linguística clássica; Linguística genética
 Lituano, 11, 72, 112; citado em, 11, 12, 13, 14, 15, 17, 18, 19, 21, 28, 45, 51, 52, 53,
 54, 56, 57, 58, 82
 Lógica, 4-5, 91, 121
 Lusaciano, 71

Macedônio, 71
 Malaio-polinésio, 78, 79
 Manchu, 78
 Manx, 69
 Mari ocidental, 77
 Mari, 77
 Messápico, 73
 Meta-gramática, 134
 Metalinguagem, 132-34
 Metátese, 50
 Método estrutural, 124
 Moabita, 73
 Modo, 109, 111
 Modulação, 95, 108-9, 112
 Møller, Herman (1850-1923), 79, 126-27
 Mongol, 77-78
 Mon-khmer, 78, 79
 Mordoviano, 77
 Mudança de palavras, 130
 Mudança de pronúncia, 130-31
 Mudança linguística, 96, 122-31
 Mudança sonora: germânico, 19-20, 24, 87; alto-alemão, 24, 87
 Mudanças de letras, 87
 Munda, 78, 79

Neerlandês, 69
 Neologismo, 66-67
 Nilótico, 79
 Nivelamento, 51-54
 Nórdico antigo, 11, 22, 70; citado em, 12, 14, 15, 17, 18, 20, 23, 51-52
 Nórdico. *Veja* escandinavo
 Norueguês, 59, 68-69; citado em, 54
 Nostrático, 80
 Número, 95, 109, 112
 Nynorsk, 69

Ob-úgrico, 76
 Oração, 98, 99
 Osco, 70
 Osseto, 72
 Ostyak, 76
 Otomano, 77

Pálavi, 72
 Palavra, 31, 32, 91, 110
 Palavras de empréstimo, 59-65, 82, 129
 Palavras estrangeiras, 60. *Veja também* Palavras emprestadas
 Páli, 72
 Paradigma, 33, 51, 53-54, 55, 133
 Pastó, 72
 Pedersen, Holger (1867-1953), 80
 Persa antigo, 72
 Polábio, 72
 Polonês, 72; citado em, 18
 Português, 70; citado em, 49

- Prácrito, 72
 Predição, 89
 Protogermânico, 22
 Proto-romance, 22
 Provençal, 70; citado em, 66
 Provençal, 70; citado em, 66
 Prussiano antigo, 71
 Psicologia, 4-5, 44, 46, 55, 121, 131
- “Riksmål,” Norueguês, 68-69
 Raiz dos elementos, 107-108
 Rask, Rasmus (1787-1832), 19, 23, 87
 Reconstrução, 83
 Regência, 107
 Relação genética, 6-7, 9, 10-31, 78-79, 82-83, 94, 95, 127, 130, 137
 Relação tipológica, 9, 91-96, 137
 Relação, 32-33, 35; implicativa, 97; não-implicativa, 97-98; implicação recíproca, 98-99; unilateral implicativa, 98-99, 134, 136-37
 Relação. *Veja* Função; genética r.; tipológica r.
 Representação dos elementos, 40-41, 48, 86-87, 113, 117, 134
 Reto-romance, 71
 Romance, 22, 46, 63, 70-71, 72
 Romênia, 72
 Romeno, 22. 71; citado em, 22
 Russo, 59, 72, 119; citado em, 18 38, 58 59
- Sagas islandesas: línguas das, 11
 Sakico, 72
 Samnita, 70
 Samnorsk, 69
 Samoieda, 76
 Sânscrito, 72. *Veja também* Índico antigo
 Santali, 78
 Sapir, Edward (1884-1939), 93
 Sardenho, 70
 Saussure, Ferdinand de (1857-1915), 123-27
 Saxão antigo, 24
 Schlegel, Friedrich von (1772-1829), 93
 Semântica, 120-21, 134. *Veja também* significado
 Semítico, 64, 73-76
 Sérvio, 71
 Servo-croata, 71
 Siamês, 78
 Siciliano, 70
 Significado, 87-88, 95, 100, 120-21, 130, 134
 Signo mínimo, 88, 91
 Signo, 31, 32, 34-35, 37-40, 45, 88, 91, 102-3, 132-33
 Signos: transformação funcional, 45-59; ação, 51-56; abreviação, 56-59; introdução de novos, 59-67
 Sílabas, 34-37, 47-48, 98
 Simbolismo-sonoro, 43
 Sino-austríaco, 6, 78, 79, 80
 Sino-tibetano, 79
 Sintagma mínimo, 110
 Sintagma, 110, 111

Síntese, 8
 Sistema de elementos, 35-37, 39; ação, 45-51; tipos, 107-8; espécies 109;
 Sistema, 123
 Sociologia, 4-5
 Sogdiano, 72
 Sons da fala, 4, 13, 41, 43, 87, 95, 113, 115, 118-19, 125, 134
 Sprachgefühl, 44, 45, 47, 54, 55, 60, 61
 Subfamília, 21-26
 Subtipos de elementos, 110-11
 subtipos, 110-11
 Sueco, 25, 54, 59, 68
 Sumério, 81

Tabu, 65-66
 Tailandês, 78
 Tarski, Alfred, 104-5
 Tcheco, 38, 72; citado em, 38, 116
 Tempos verbais, 109, 111
 Tendências, 130-31
 Texto, 97, 133
 Tai-chinês, 78, 79, 93
 Thomsen, Vilhelm (1842-1927), 65, 77
 Tibeto-birmanês, 78, 79
 Tipo aglutinante, 92-93
 Tipo isolante, 91-92, 93, 94, 108
 Tipo polissintético, 92-94
 Tipologia das línguas, 9, 91-96, 121 130-31, 136
 Tocário, 11, 73; citado em, 11, 12, 13
 Tonicidade, 47
 Trácio, 73
 Tradutibilidade, 135
 Transformação dos signos, funcional, 45-59
 Transformação funcional dos signos, 45-59
 Tungus, 78
 Turco antigo, 77
 Turco, 63, 77, 94, 110; citado em, 92
 Túrquico, 77

Ucraniano, 72
 Udmúrtia, 77
 Úgrico, 76
 Umbro, 70
 Umbro-Samnita, 70
 Unidades, 95-96
 Uniformes, 101
 Urálico, 76-77, 80, 92, 93
 Uso linguístico, 31, 32-44, 54, 57 5g 60-61, 91,95,96, 115-21, 122, 130-31, 132

Variação, 113-14
 Variante, 58, 113-14, 117-19
 Variável, 40
 Variedade localizada, 114
 Variedade, 113-14, 128
 Védico, 72. *Veja também* Índico antigo
 Venetie, 73

Verner, Karl (1846-96), 20, 24
Vestimentas populares, 101
Vestimentas, 101
Vietnamita, 78, 79
Vocoide, 110, 115-16, 118-19
Vogal, 27, 35, 95, 98-100, 110-12, 116
Vogul, 76
Votyak, 77

Wends, 71
Wulff, Kurt (1881-1939), 79

Yaghnobi, 72

Zaratustra, 72
Zyriano, 77

φ, 15
*, 16

3 ESTUDO CONTRASTIVO DA IMANÊNCIA NOS *PROLEGÔMENOS A UMA TEORIA DA LINGUAGEM* E EM *A LINGUAGEM*

A partir do que foi elaborado até o momento, delegamos a este capítulo a tarefa de discorrer, sem o intuito de ser exaustivo, sobre a construção da imanência em *A linguagem*. Além de formar o núcleo de nosso estudo complementar à tradução, a imanência é um conceito central não somente nesta pesquisa, mas também dentro do quadro teórico hjelmsleviano, pois sua constância demonstra a coerência interna do trabalho de Hjelmslev e serve como fio condutor entre todas as obras, tanto para as que possuem um teor mais histórico quanto para aquelas mais teóricas. Dentro desse espectro, encontramos a persistência desse conceito como o núcleo gerador da glossemática.

Podemos observar a questão da imanência no quadro teórico dos estudos hjelmslevianos a partir de estudos recentes como os de Badir (2014), Zinna (2014) e Beividas (2015). Veremos em detalhe como esse conceito se articula mais à frente, pois devemos, antes, situar a questão epistemológica em Hjelmslev, visto que é a partir da abordagem desse ponto que se desenvolve a perspectiva de uma imanência conciliadora, nas palavras de Badir (2014). O pensamento hjelmsleviano propõe um recorte preciso no panorama das principais discussões filosóficas e de método que ocorreram – e ainda ocorrem – no ocidente. Em princípio, a partir de dois direcionamentos construídos com intenções distintas que guiaram nossa proposta de tradução, pode-se elaborar uma primeira perspectiva sobre esse tópico. Nos *Prolegômenos a uma teoria da linguagem* (2013 [1943]), elabora-se um progressivo avanço em direção à virada epistemológica que alça a imanência – da teoria – a um patamar de regência em relação às chamadas atitudes transcendentais e, também, à própria imanência do método. Essa ascensão da imanência torna-se princípio regulador de uma terceira via do conhecimento, como é debatido por Beividas (2015). Esse pensamento atua em nosso estudo como suporte articulador para pensarmos a imanência nos *PTL*.

Dessa base epistemológica construída por Hjelmslev – apesar de evitar estender-se sobre a discussão e deixá-la a cargo dos epistemólogos –, o teórico elabora o princípio do empirismo, do qual a teoria da linguagem, enquanto operação analítica, depende inteiramente. Tal princípio elabora-se sobre as seguintes formulações: a teoria da linguagem deve ser não contraditória, exaustiva e simples; a primeira formulação prevalece sobre a segunda e a segunda sobre a terceira (2013 p.11). Em Badir (2014), tem-se a proposição de que o princípio do empirismo proposto por Hjelmslev é, na

verdade, uma articulação possível de uma mesma categoria. Para entender isso, Badir apresenta o empirismo como uma categoria que não é definitiva no pensamento de Hjelmslev, e dela deduzem-se outras categorias como a simplicidade e a economia, dentre outras que abordaremos adiante. Por essa abertura, considera-se o empirismo como um princípio lacônico, o que oferece margem para múltiplas interpretações sobre sua constituição. Assim, propõe-se que o princípio da adequação e o princípio da imanência são manifestações de uma mesma sistemática, incluindo aqui o princípio do empirismo. A adequação atua como um princípio regulador do empirismo, pois é a partir dela que a teoria é empírica. Em outras palavras, “certas premissas enunciadas na teoria preenchem as condições necessárias para que esta se aplique a certos dados da experiência” (Hjelmslev, 2013, p. 16). Do outro lado, a imanência da teoria estabelece-se a partir de um cálculo arbitrário e, em si mesma, não depende de qualquer dado da experiência (2013, p.16).

Ainda segundo Badir, o princípio do empirismo não se esgota em si mesmo, mas se analisa em outros princípios: da simplicidade, da economia, da redução, da generalização e, por fim, o princípio da descrição exaustiva, que é derivado do princípio do empirismo e do princípio da redução (Badir, 2014, p. 73). Como vimos anteriormente, a simplicidade, derivada da análise do princípio do empirismo, é a única maneira pela qual se pode afirmar se determinada ocorrência não contraditória e exaustiva é preferível em comparação com outra que também o é. Ela atua como a *navalha de occam* do princípio do empirismo e regula a extensão final da análise empreendida pela teoria, pois, entre duas análises exaustivas, por exemplo, o teórico deve admitir aquela que seja mais simples. Desse primeiro nível derivacional, da simplicidade, emanam os princípios da economia e o da redução. A cada passo da análise, não há como prever que aquele será o último nível e, por isso, a teoria deve sempre avançar na análise em busca das menores grandezas possíveis. No entanto, se a análise demonstrar que o nível posterior produziu um resultado menos simples que o nível anterior, deve-se deter a descrição e retornar à solução mais simples, segundo o princípio da economia (Hjelmslev, 2013, p. 65). A redução articula-se ao princípio da economia ao estipular que a análise deve continuar até que se esgotem as possibilidades de continuidade e se atinja, portanto, o nível em que se registre o menor número de objetos que for possível. O objetivo da descrição sempre é que se chegue ao registro do menor número de elementos possível (2013, p. 66)

A adequação entra nessa função como uma exigência da teoria, que é empírica, por ser adequada a determinado objeto - a linguagem. Nesse caso, a teoria da linguagem

entra em uma espécie de ciclo, pois ela própria define e é definida por seu objeto: uma teoria da linguagem não teria sentido sem um objeto “linguagem”; a exigência desse objeto se torna então uma necessidade prática (Badir, 2014, p. 77). Se assim considerarmos a adequação, devemos também compará-la ao que Hjelmslev nos diz para introduzir o princípio do empirismo: “uma teoria, para ser a mais simples possível, só deve elaborar a partir das premissas que sejam necessariamente exigidas por seu objeto” (2013, p. 11). As premissas exigidas pelo objeto regulam e adequam a teoria.

À imanência, como terceira manifestação dessa mesma sistemática, compete a tarefa de conciliar contrários epistemológicos: arbitrariedade e adequação (Badir, 2014, p. 78). Essa atitude conciliatória registra-se em torno de uma concepção que visa permitir uma visão global do objeto do conhecimento – atitude essa que assinala, em última instância, a regência do princípio da imanência sobre a transcendência.

Badir afirma que a primeira problemática a ser resolvida, em relação às proposições de Hjelmslev, diz respeito a essa conciliação entre um sistema dedutivo puro – a teoria – e a ciência empírica – arbitrariedade e adequação. Essa questão é resolvida pelo teórico dinamarquês na medida em que o vínculo da teoria da linguagem e o conhecimento empírico baseia-se no princípio da adequação o que culmina, entretanto, em uma concepção, segundo Badir, um tanto aproximada do convencionalismo de Poincaré: a aplicabilidade. Dessa maneira, a teoria, pela sua arbitrariedade, não indica a existência de uma determinada aplicação, porque, caso o fizesse, estaria lidando diretamente com os dados de uma suposta experiência não analisada, uma substância linguística. No entanto, só conhecemos, ou supomos conhecer, tal substância a partir da forma linguística. Então, só se reconhece uma semiótica, enquanto plano analisado pela teoria, na medida em que ela é submetida à análise em um plano do conteúdo e um plano da expressão. Esses dados da experiência só respondem à teoria enquanto forma linguística. Hjelmslev (2013, p. 16) nos diz, então, que o teórico sabe, por experiência, que certas premissas estabelecidas pela teoria, preenchem as condições necessárias para que esta seja aplicável a determinados dados da experiência, ou seja, adeque-se a eles. Esses supostos dados da experiência só se confirmam como aplicáveis pela teoria quando submetidos à análise. Dessas formulações, tem-se que o objeto da teoria da linguagem é o texto, mas o que há antes da análise são apenas *supostos textos*. Os dados da experiência para Hjelmslev (2013, p. 14) constituem uma totalidade absoluta não analisada, isto é, são apenas supostos, pois não entraram na teoria. Enquanto totalidade não analisada e

apenas suposta, não há nada que possa ser dito sobre esses dados, a não ser supor sua existência e, inversamente, tais dados não afetam de maneira alguma a teoria.

O que culmina dessas articulações metateóricas é uma proposição de nível epistemológico que atualiza o devir saussuriano do objeto: conhecer e descrever um objeto é, em suma, um mesmo ato. Essa constituição e articulação entre epistemologia e método surge nos *PTL*, mais especificamente no capítulo 15, “Esquema e uso linguístico”, em que Hjelmslev formula a tomada de conhecimento da linguagem como um ato puramente descritivo:

A tarefa principal do linguista, portanto é construir uma ciência da expressão e uma ciência do conteúdo sobre bases internas e funcionais, sem admitir dados fonéticos ou fenomenológicos na ciência da expressão, nem dados ontológicos ou fenomenológicos na ciência do conteúdo (o que não significa, evidentemente, que se irá negligenciar as premissas epistemológicas sobre as quais toda ciência se baseia) (Hjelmslev, 2013, p. 81).

Em *A linguagem*, já no capítulo inicial, “A função linguística”, temos uma formulação que possui o mesmo intuito articulatório:

Atingimos discernimento, familiarizamo-nos, entendemos ou compreendemos a linguagem da mesma forma que assimilamos outros objetos – por meio de uma descrição. E descrever um dado objeto pode significar apenas uma coisa: dar conta das relações das quais faz parte ou daquelas que o constituem. A essas relações, ou dependências, registradas por uma descrição científica, daremos o nome de *funções*.²⁷

3.1 Da imanência à *Imanência*: o desenvolvimento teórico nos *PTL* e em *A linguagem*

Propomos, então, uma leitura comparativa que tem como objetivo primário abrir margem para a exploração desse outro caminho do pensamento hjelmsleviano. Como já apontamos, *A linguagem* (1970) possui uma elaboração teórica que parece fluir, propositadamente, em direção a um texto que vise um público menos acadêmico, em oposição ao texto denso e estritamente dimensionado dos *Prolegômenos a uma teoria da linguagem* (2013). Essa característica já é apontada na introdução escrita por Whitfield para a edição em inglês. O renomado tradutor nos diz que, diferentemente do que acontece

²⁷ “We gain insight into language, acquaintance with it, understanding or comprehension of it in the same way as we gain insight into other objects – through a description. And to describe an object can mean only one thing: to give an account of the relationships into which it enters or which enter into it. Such relationships, or dependences, registered by a scientific description, we shall call *functions*.” (Hjelmslev, 1970, p. 8, tradução nossa).

nos *PTL*, no livro com foco mais popular – isto é, *Language* – tem-se como objeto principal as relações genéticas entre as línguas e como essas relações, genéticas ou não, estabelecem-se. Nos *PTL*, observam-se os contornos da teoria glossemática voltada para um público que possua uma formação linguística. Por isso, como já dissemos anteriormente, notamos uma construção mais enxuta e direcionada – mais *definicional*. Enquanto isso, em *A linguagem*, o texto abre mão de certas escolhas e termos mais específicos para, supostamente, tornar a leitura menos carregada e mais simples. Assim, iniciamos nosso percurso que visa compreender o desenvolvimento desses dois métodos em relação à imanência.

Podemos observar esse movimento ao compararmos como Hjelmslev introduz e delimita as definições de *funtivo* e *função*, especialmente nas escolhas lexicais que são feitas. Nos *PTL*, o teórico deixa claro que um funtivo não *possui* uma função, mas a contrai (2013, p. 39). Entendemos que esse movimento é feito pois, caso contrário, pressuporia a existência prévia do funtivo antes mesmo da contração funcional, perspectiva essa que é rechaçada por Hjelmslev e que se alinha ao que ele chama de *realismo ingênuo*:

Os relacionamentos ou as dependências que o realismo ingênuo considera secundários e como pressupostos dos objetos tornam-se, para nós, essenciais: são a condição necessária para que existam pontos de intersecção (2013, p. 28).

Na perspectiva desse realismo ingênuo, há uma existência prévia dos objetos analisados no que concerne aos relacionamentos estabelecidos, ou seja, tais relações não são a razão para a existência dos objetos, pelo contrário, são características secundárias. Entretanto, para Hjelmslev, o objetivo da análise não é executar uma mera divisão, mas sim reconhecer as funções que existem para, então, definir os pontos de intersecção entre elas – os funtivos.

Será reconhecido, portanto, sem dificuldades, que no fundo o essencial não é dividir um objeto em partes, mas sim adaptar a análise de modo que ela seja conforme às dependências mútuas que existem entre essas partes, permitindo-nos prestar contas dessas dependências de modo satisfatório. (2013, p. 28)

Portanto, Hjelmslev nos *PTL*, texto estritamente mais diretivo e objetivo, não pretende deixar margem para uma interpretação que se desvia dos princípios da glossemática. Esse direcionamento objetivo e de contornos lexicais bem definidos não

tem a mesma força em *A linguagem*, cuja elaboração teórica parece deslizar, propositalmente, em direção a um texto que vise um público menos acadêmico.

Para estabelecermos uma base para comparação, apresentaremos a seguir como as definições para funtivo e função são estabelecidas em *A linguagem*. Dessa análise prévia, observaremos que em *A linguagem* propõe-se elaborar uma glossemática que não está no nível mais superficial do texto, ou seja, que não se deixa perceber com facilidade, mas sim por seus contornos teóricos e metateóricos subjacentes à manifestação textual. Dessa forma, é a partir da ausência do termo basilar – glossemática – constatada por Whitfield, bem como de nossa constatação da ausência (apenas) aparente, das formas mínimas que a teoria estabelece em sua base explicativa, que formulamos os alicerces da comparação entre a construção dessa glossemática prevista para o público inexperto e a glossemática teórica e metateórica.

3.1.1 *Function, fonction, funzione, función, funktion, função.*

Antes de tudo, é necessário apresentarmos brevemente como o conceito de *função* é organizado na teoria hjelmsleviana e, conseqüentemente, o conceito de *funtivo*. Nos *PTL*, à função é dado o estatuto de uma dependência que preenche as condições de uma análise – “uma passagem dedutiva de classe para componente e componente de componente” (2013, p. 27). Hjelmslev nos diz, entretanto, que essa passagem da classe ao componente não deve ser reduzida simplesmente a uma decupagem, mas deve adaptar-se “às dependências mútuas que existem entre essas partes” (p. 28). Esse procedimento assegura a adequação da análise e, ainda, torna as relações funcionais o foco do proceder linguístico o que, em última instância, produz as seguintes postulações: a totalidade do objeto posto em análise depende unicamente da soma das funções contraídas, sejam elas em relação à hierarquia – a totalidade -, em relação às funções de mesmo nível – mesma classe -, ou em relação às funções do grau seguinte – dos componentes (p. 28). O fato de que esses relacionamentos assumem posição primária dentro da análise da teoria glossemática torna os pontos de intersecção entre eles secundários. A esses pontos, dá-se o nome de *funtivos*. Entende-se por funtivo um “objeto que tem uma função em relação a outros objetos” (p. 39). Essa definição possui um caráter puramente relacional, ou seja, um funtivo só recebe essa classificação pelas funções contraídas. Desse modo, tem-se que funções também podem ser funtivos, pois existem funções entre funções. Por fim, àqueles funtivos que não forem também funções, será dado o título de *grandeza* (p. 39). Não nos

estenderemos aqui sobre os tipos de funções e suas derivativas, pois seria uma digressão além do nosso objetivo. Aqui, devemos notar que o Hjelmslev dos *PTL* se vale das duas formas para expressar a função: a partir do ponto de vista possessivo – “tem uma função” – e a partir do ponto de vista contrativo – “Uma classe que contrai uma função com uma ou várias classes no interior de uma mesma série será chamada de soma” (2013, p. 85) – movimento que não ocorre em *A linguagem*, cuja construção apoia-se inteiramente sobre a primeira formulação.

Para nossa análise, interessa agora que vejamos como o conceito de *função*, especificamente em *A linguagem*, permite que observemos o que afirmamos anteriormente, ou seja, que a escolha de algumas formulações teóricas parece balizar em direção a um texto que vise um público menos acadêmico. Para tanto, abaixo (Tabela 1) organizam-se formulações encontradas do termo *função* nas edições em inglês, francês, italiano, espanhol, dinamarquês e português – nossa tradução – e nos ajudará a entender como essa definição é abordada no livro em questão:

Tabela 1 – Apresentação de uma “função” em *A linguagem*

Idioma	Ocorrência	Verbos
Inglês	“it is the one <i>linguistic structure</i> that has function to the other” (1970, p. 31)	to have - has
Francês	“c’est une <i>structure linguistique</i> qui a une fonction par rapport à l’autre. (1966, p. 54)	avoir - a
Italiano	“è una <i>struttura linguistica</i> che ha una funzione rispetto all'altra.” (1963, p. 35)	avere - ha
Espanhol	“es una <i>estructura lingüística</i> que tiene una función con relación a la otra” (1971, p. 40)	tener - tiene
Dinamarquês	“det er den ene <i>sprogbygning</i> , der har funktion til den anden” (1973, p. 34)	at have - har
Português	“é uma <i>estrutura linguística</i> que tem função com a outra.”	ter - tem

Fonte: elaborada pelo autor.

Para a tabela acima, devemos considerar a *estrutura linguística* como um dos fustivos da função que é estabelecida. Assim sendo, em resumo, ambas as estruturas só podem ser consideradas como tais a partir da função que é mantida entre elas. Porém, nosso foco é comparar a apresentação dessa função com a definição que é dada nos *PTL*.

Enquanto em *A linguagem* é dito que um funtivo tem/possui uma função com outro funtivo, nos *PTL*, Hjelmslev conceitua esse processo como “um funtivo que *contrai* sua função” (p. 39). A diferença consiste, efetivamente, no valor formal dessa definição. Por um lado, “ter/possuir” não é mais do que uma escolha lexical, ou puramente *operacional*, que em *A linguagem* observamos em outras passagens do livro como em: “[...] como uma função entre línguas formado pelo fato de que categorias em cada língua têm uma função com categorias em cada uma das outras línguas”²⁸, na qual observamos o mesmo verbo destacado na tabela acima. Hjelmslev define essa categoria operacional como tendo duas possibilidades dentro da teoria: “[...] definições que, num estágio mais avançado, se transformarão em definições formais e, de outro lado, de definições puramente operacionais cujos conceitos definidos não estarão no sistema de definições formais.” (2013, p. 25). Acreditamos que a ocorrência em questão seja mais bem entendida a partir da primeira opção, visto que ela se torna uma definição propriamente estabelecida no sistema da teoria que é apresentado nos *PTL* e no *Résumé d'une théorie du langage* (1973). No primeiro caso, temos a definição 10 que diz o seguinte sobre a contração: “diz-se de um funtivo: *contrair* sua função” (2013, p. 137). No segundo, mais detalhado, sobre o funtivo, Hjelmslev nos diz que “Um *funtivo* (símbolos p, q, r...) é um objeto que tem Função com outros objetos. É dito que um funtivo *tem função com* (e não “é função de”) outro funtivo. Diremos de um funtivo que ele *contrai* sua função.”²⁹ (2014, p. 6). Assim, vemos que nos *PTL* e no *Résumé* a escolha da palavra *contrair* envolve, na verdade, uma *operação formal*, ou seja, temos de fato uma definição formal que faz parte do sistema da teoria. No *Résumé*, podemos notar, inclusive, a passagem de uma definição operacional para uma definição formal quando “*avoir une fonction à*” passa a “*contracte sa fonction*”.

Em suma, podemos perceber que, embora exista uma confluência de ideias que contribuem para o mesmo objetivo em ambos os livros, esse objetivo é alcançado de maneira distinta. Em *A linguagem*, temos um menor comprometimento com a precisão da metalinguagem utilizada, visto que ela, estabelecida por conceitos operacionais, que se assentam no uso mais corrente da linguagem. Os conceitos operacionais podem ser modificados posteriormente e tornados definições formais – ou ainda eliminados do

²⁸ N. T.: Tradução do original inglês: “[...] as a function between languages consisting in the fact that categories in each language have function to categories in each of the others”.

²⁹ N. T.: Tradução do original francês: “Un Fonctif (symbole: p, q, r...) est un objet qui a une fonction par rapport à d’autres objets. – Un fonctif est dit *avoir une fonction à* (et non “être fonction de”) un autre fonctif. On dira d’un fonctif qu’il *contracte sa fonction*.”

sistema de definições formais, como vimos mais acima. Essa escolha permite uma leitura também menos comprometida com a totalidade da teoria, uma vez que as definições formais estabelecem uma rede intrincada de interdefinições de difícil apreensão. Isso sugere o direcionamento a um público menos experiente em matéria linguística, como apontado por Whitfield. Esse menor rigor definicional contribui para que visualizemos também o apagamento discursivo da imanência metodológica e epistemológica e o impacto que isso gera na argumentação desenvolvida em cada livro.

3.1.2 A imanência nos PTL

Podem-se dividir os capítulos dos *PTL* em três grupos segundo um critério de conteúdo apresentado: o primeiro (capítulos 1-9) destinado aos critérios gerais da teoria da linguagem; o segundo (capítulos 10-20) voltado às especificidades da teoria linguística e, por fim, o terceiro (capítulos 21-23) dedicado às relações entre linguagem e não-linguagem. Além disso, de forma quantitativa, podemos elencar onze ocorrências do vocábulo *imanência* e suas derivações durante a extensão dos *PTL*. Nosso intuito será, portanto, duplo: produzir um comentário geral sobre a teoria da linguagem e sua apresentação – e como ela traz consigo o caráter complexo da imanência desde os nove capítulos iniciais, ou seja, desde a apresentação dos critérios gerais – e, conjuntamente, analisar a apresentação dessas ocorrências – da imanência – ao longo do livro. Como já salientamos, a linguagem apresentada por Hjelmslev é de extrema concisão e cuidado, sempre apoiada em uma rigorosa terminologia, sobre a qual nos detivemos brevemente no capítulo anterior, o que reflete de forma acentuada a decisão de produzir um livro mais academicista em relação ao que encontramos em *A linguagem*.

Para nosso comentário inicial, utilizaremos os elementos do quadro tensivo de Zilberberg (2011) e Fontanille e Zilberberg (2001) como aparato teórico para entender os processos de estruturação discursiva da imanência nos *Prolegômenos a uma teoria da linguagem* (2013). Especificamente, para analisar o processo construído na sintaxe tensiva da atualização dos valores atribuídos à imanência, buscaremos suporte no encadeamento dos modos de existência - virtualização, atualização, potencialização e realização – os quais relacionam-se aos sujeito e objeto narrativos. Tais modalizações atuam tanto sobre o sujeito do ser, atribuindo existência modal ao sujeito de estado, como sobre o sujeito do fazer, sendo a responsável pela construção da competência modal deste último.

Segundo Greimas e Courtés (2008), a existência virtual vincula-se ao eixo paradigmático, pois relaciona-se ao sistema pressuposto e, por conseguinte, é manifestada por uma existência *in absentia*, ou seja, marcada pela ausência. Assim, a passagem do primeiro nível virtual para o segundo, atualizado, é marcado também pela passagem do sistema para o processo. Desse segundo nível, chega-se a um terceiro, caracterizado pela manifestação discursiva na sintagmática. Esses processos, segundo Prado (2013), constituem-se pela relação junctiva, pois sujeito e objeto, enquanto não tomados por uma relação que os correlacionem, permanecem apenas na virtualidade. O último nível, da potencialização, foi concebido por Fontanille e Greimas com base na *não-conjunção* (1993^a, p. 52) e revisto por Fontanille e Zilberberg em *Tensão e significação* (2001):

Os modos de existência, ou *modalizações existenciais*, fornecem-nos desde já uma sintaxe canônica, que cruza dois percursos, como no quadrado semiótico: a inatividade (a potencialização) constitui uma “perda” de densidade existencial, provocada pela anulação do foco, perda que conduz da presença (realizante) à ausência (virtualizante); inversamente, a perda (atualizante) proporciona um ganho de densidade existencial, devido à intensidade do foco, no caminho que leva da ausência à presença (Fontanille; Zilberberg, 2001, p. 135).

Portanto, a partir das formulações de Greimas (1973), Zilberberg (2011) – para as questões sobre o andamento apontados anteriormente em nossa introdução – e Beividas (2015) – para o surgimento brusco da imanência epistemológica –, buscamos entender como esse processo ocorre nos *PTL* e, paralelamente, como ele se desenvolve em *A linguagem*, no capítulo seguinte.

3.1.2.1 A teoria da linguagem e a imanência

Acreditamos que, já a partir das formulações iniciais do enunciador dos *PTL*, temos o estabelecimento de uma lógica implicativa desde o apontamento em direção a uma linguística de nível epistemológico e, conseqüentemente, uma imanência que também o seja (Hjelmslev, 2013, p. 4). Essa percepção se opõe à de Beividas (2015), especificamente em relação ao surgimento “brusco” da imanência de nível epistemológico na construção dos *PTL*, a qual viria à tona apenas no último capítulo. Nossa proposta entende que a construção dessa imanência perpassa todos os capítulos. Nosso intento é demonstrar, assim, uma construção gradual que torna o ato tônico do último capítulo uma ocorrência que fora sistematizada ao longo do livro e, portanto, realiza-se implicativamente a partir de nossa leitura.

Inicialmente, a linguística atua como uma força secundária, contributiva, concessiva, em relação aos estudos do conhecimento. Porém, gradativamente, por força da repetição, a recorrência da construção concessiva acaba por entrar no esquema de previsões do enunciatário e, assim, proporciona uma queda relativa no andamento visto de modo global. Vejamos, inicialmente, como essa lógica principia o seu estabelecimento:

A linguagem é o instrumento graças ao qual o homem modela seu pensamento, seus sentimentos, suas emoções, seus esforços, sua vontade e seus atos, o instrumento graças ao qual ele influencia e é influenciado, a base última e mais profunda da sociedade humana (Hjelmslev, 2013, p. 1).

A partir da descrição acima, pode-se pensar a teoria da linguagem de Hjelmslev a partir da apresentação que ele propõe para a linguagem: a linguagem é um instrumento, uma ferramenta para atingir fins exteriores a si mesma. Porém, já adiante, Hjelmslev questiona se a linguagem não é a própria fonte das manifestações da vida (2013, p. 2). Pode-se entrever nessa afirmação a apresentação da linguagem agora como uma base imanente para a transcendência. Esse questionamento aparentemente sem maiores efeitos serve como um primeiro apontamento para o que viria nos próximos capítulos em relação à própria teoria da linguagem. Em contrapartida, já na página seguinte (p. 4), Hjelmslev nos diz que a construção imanente da teoria da linguagem poderá, então, ter sua importância reconhecida por meio de sua contribuição à epistemologia geral. Essa ascensão da imanência, caracteriza-se, em nossa leitura, como o primeiro apontamento para a construção gradual de uma lógica implicativa da imanência de nível superior:

Segundo nossa perspectiva, a introdução – e repetição – dessa imanência epistemológica pode ser observada nas seguintes passagens: “[...] se ela não é (a linguagem) tudo isso: a própria fonte do desenvolvimento dessas coisas.” (2013, p. 2). Nesse trecho, podemos perceber que Hjelmslev coloca a linguagem em uma posição inversa àquela que é normalmente observada nas pesquisas com atitude transcendental, como ele mesmo cita um pouco abaixo no mesmo capítulo. Com essa afirmação, a língua deixa de ser um meio para se estudar os fenômenos físicos e fisiológicos, por exemplo, e assume a origem desses fatos. Logo em seguida, temos que: “[...] é só artificialmente que a pesquisa pode ser dirigida para o próprio meio do conhecimento” (2013, p. 3). É então um procedimento teórico imanente que permite entender a língua por ela mesma e para ela mesma, ou metalinguisticamente. Por fim, na página seguinte, Hjelmslev enuncia que

a linguística assumirá uma posição contributiva em relação às teorias gerais do conhecimento: “[...] a curto prazo é através de sua contribuição à epistemologia geral que a linguística revelará, de modo incontestado, sua importância.” (p. 2013, p. 4). Esse posicionamento ainda é um tanto desprezioso, pois a teoria da linguagem, como já apontamos anteriormente, desponta no último capítulo incorporando não só o sistema linguístico, mas também o homem, a sociedade e todo o domínio do saber humano (2013, p. 133).

Assim, o que antes se apresentava concessivamente (a teoria da linguagem como uma unidade epistemológica, metalinguística e imanente), finda por tornar-se implicativo. No capítulo 1 dos *PTL*, Hjelmslev (2013, p. 3) elabora que a linguagem deve ser apreendida a partir de um todo autorregulado, uma estrutura única, o que se segue a um tratamento científico ao qual essa unidade é submetida. Inicialmente, tem-se uma construção concessiva, pois a linguística adentra, como colaboradora, o campo epistemológico geral. Essa percepção constrói-se sob uma imanência de método que é guiada por um tratamento epistemológico que lhe é superior, uma imanência epistemológica, em nosso entendimento. Ou seja, a teoria ganha um aspecto gnosiológico, como aponta Badir:

Se a teoria da linguagem fornece a sua própria classificação, ela não espera de outra teoria essa atribuição, mesmo que esta se declare qualificada para o papel de classificar os campos do conhecimento. [...] Portanto, o que a teoria da linguagem define por si só também é válido para a organização de todos os domínios do conhecimento.³⁰

Dessa forma, a imanência linguística torna-se central na teoria da linguagem já desde o princípio da exposição pois é a partir dela que todos os desenvolvimentos posteriores da teoria surgirão, o que vai contra os chamados estudos transcendentais (2013, p. 4). Os apontamentos feitos acima, em que vemos um direcionamento para uma atitude epistemológica, centram-se em sua maioria, em relação à linguagem e não, ainda, à teoria da linguagem. É somente a partir da sugestão de que a linguística poderá contribuir em um nível das ciências gerais (2013, p. 4) que podemos perceber a primeira ocorrência desse processo em direção a uma teoria da linguagem de nível epistemológico e, conseqüentemente, à imanência epistemológica. Assim, tais elementos inicialmente

³⁰ “Si la teoría del lenguaje proporciona su propia clasificación, no espera de otra teoría que le asigne una, aun cuando incluso ésta se declarara habilitada para el rol de clasificar los conocimientos. [...] Por lo tanto, lo que la teoría del lenguaje define por sí misma es también válido para la organización de los saberes en su conjunto” (Badir, 2015, p. 81, tradução nossa).

não previstos, gradativamente, são trabalhados em direção à proposta de Hjelmslev. É nesses termos que podem se tornar implicativos, em um segundo momento, uma vez que esse esquema entra nas previsões do enunciatário.

O segundo capítulo dos *PTL* trata de alçar a estrutura linguística imanente a um papel imperativo. A imanência da linguagem adquire contornos dominantes: agora é a organização do sistema linguístico que rege os outros sistemas (físicos, psicológicos, lógicos, etc.), que se estabelecem somente em função da estrutura linguística (2013, p. 8). Essa centralidade da estrutura linguística se dá fundamentalmente pela abordagem que toma a língua como um fim em si mesma. A partir dessa afirmação, a língua tem sua imanência garantida pois, para que possamos acessar outros sistemas – ou semióticas – havemos antes de passar pelo crivo da língua. Em outras palavras, o estudo da língua deve ser imanente pois essa é uma exigência do próprio objeto – a língua.

O terceiro capítulo dedica-se à introdução do princípio do empirismo, o qual, como citamos acima, relaciona-se estreitamente ao princípio da adequação e, como propõe Badir (2015), são manifestações de uma mesma sistemática observada a partir de diferentes pontos de vista, assim como o princípio da imanência. A língua, que antes se submetera a um tratamento científico (HJELMSLEV, 2013, p. 3), agora exige suas próprias premissas (p. 11), que passam a ser elaboradas, logicamente, em torno da organização interna do objeto analisado, o que as torna, em última instância, parte integrante do próprio objeto. Essa abordagem apontada nos dois capítulos citados indica o que chamamos de imanência metodológica, mas que já apresenta um encaminhamento para a imanência epistemológica. Podemos definir essa progressão da seguinte maneira: à primeira, alinham-se características que a tornam *secundária* em relação ao seu papel dentro da teoria, ou seja, a ela é dado o papel de “[...] ser submetida a um tratamento científico [...]” (2013, p. 3). Para a segunda, já se observa nesses mesmos capítulos iniciais uma progressão rumo a um papel *primário*, o qual poderíamos chamar também de agentivo, no sentido de executar uma ação ao invés de recebê-la. Isso pode ser observado quando Hjelmslev trata da contribuição que o estudo da estrutura da língua pode dar à epistemologia (2013, p. 4). Em primeiro lugar, entende-se a imanência metodológica a partir de uma atitude centralizada na própria estrutura linguística, ou seja, ao linguista cabe desenvolver uma teoria da linguagem por meio de definições que sejam derivadas apenas da própria estrutura linguística (Bevidas, 2015, p. 5). Desse modo, deve-se evitar qualquer atitude transcendental que permita que a língua seja vista a partir de pontos de vista físicos, fisiológicos, psicológicos ou lógicos (Hjelmslev, 2013, p. 2). Em resumo,

podemos tratar de dois modos de construir a imanência: o primeiro que se constrói sobre formulações que dão à imanência características as quais definimos como secundárias, ou seja, que tem seu papel limitado apenas ao da estrutura linguística analisada. O segundo modo é o da imanência epistemológica, primário, no qual ela tem um papel ativo desde a contribuição que a análise da estrutura linguística pode ter para a epistemologia geral até a formação de novas bases para o estudo linguístico. A tabela abaixo organiza pelo menos uma ocorrência da imanência metodológica que localizamos em nossa análise de cada um dos nove capítulos iniciais:

Tabela 2 – Ocorrências da imanência como método nos *Prolegômenos*

Imanência metodológica

“É só deste modo que a língua enquanto tal poderá ser submetida a um tratamento científico [...]” (2013, p. 3, grifos nossos)	Cap. 1
“Uma teoria que procura a estrutura específica da linguagem com a ajuda de um sistema de premissas exclusivamente formais [...]” (2013, p. 7, grifos nossos)	Cap. 2
“O objetivo da teoria da linguagem é verificar a tese da existência de um sistema subjacente ao processo, e a tese de uma constância que subentende as flutuações [...]” (2013, p. 9, grifo nosso)	Cap. 2
“Uma teoria para ser a mais simples possível, só deve elaborar a partir das premissas que sejam necessariamente exigidas por seu objeto ” (2013, p. 11, grifos nossos)	Cap. 3
“O único procedimento possível para o isolar o sistema que esse texto subentende é uma análise que considera o texto como uma classe analisável em componentes” (2013, p. 14, grifos nossos)	Cap. 4
“Com a terminologia que escolhemos pudemos caracterizar o método da teoria da linguagem como sendo necessariamente empírico e dedutivo [...]” (2013, p. 15, grifo nosso)	Cap. 5
“[...] uma teoria [...] tem por objetivo elaborar um procedimento por meio do qual se possa descrever, não contraditoriamente e exaustivamente, objetos dados de uma suposta natureza.” (2013, p. 19, grifo nosso)	Cap. 6
“Procurando uma constância no próprio interior da língua e não fora dela (cf. Cap. 2), a teoria procede inicialmente a uma limitação necessária, mas apenas provisória, de seu objeto” (2013, p. 23, grifo nosso)	Cap. 7
“A teoria da linguagem, cuja tarefa principal é explicitar – remontando o mais longe possível – as premissas científicas da linguística [...]” (2013, p. 25)	Cap. 8

“[...] o **objeto** examinado **só existe em virtude desses relacionamentos** ou dessas dependências; [...]” (2013, p. 28, grifo nosso) Cap. 9

Fonte: elaborada pelo autor.

Os trechos acima listados representam exemplos que demonstram a abrangência da imanência de escopo metodológico. Pode-se perceber pelas passagens destacadas, que o método de análise da teoria da linguagem tem compromisso somente até os limites da estrutura linguística. O tratamento científico ao qual a língua é submetida no primeiro e segundo exemplos advém da teoria da linguagem e seu objeto, sendo a língua, possui extensão limitada à estrutura específica dessa mesma língua. Seguindo esse raciocínio, verificamos o mesmo desenrolar nos exemplos que vêm a seguir: a teoria da linguagem verifica a existência de um sistema subjacente a um processo; só se constrói a partir das premissas exigidas por seu objeto (a língua); é formada por uma análise que considera o texto como uma classe decomponível em componentes; define-se a partir de sua metodologia como um procedimento empírico e dedutivo; tem como objetivo elaborar um procedimento que possa analisar os objetos de uma mesma dada natureza (linguística); procura uma constância – o sistema – no interior da própria língua; busca explicitar as bases científicas da linguística; e considera que o objeto analisado – a língua – só existe em função dos relacionamentos que contrai em seu interior.

Vejamos agora como a imanência epistemológica se faz presente nesse mesmo contexto. a teoria da linguagem, como um sistema dedutivo puro, não depende em nada dos dados da experiência para validar suas implicações (Hjelmslev, 2013, p. 16). Essa construção, arbitrária em relação à experiência e adequada em relação à aplicabilidade, fundamenta-se sobre a ideia de que para que se tenha uma teoria da linguagem tão geral quanto possível, é necessário que sejam elaborados princípios tão largos que acabam por transpor os limites entre a teoria da linguagem e outras teorias. Ou seja, princípios esses que seriam aplicáveis também a outras linguagens e semióticas. Essa posição apresentada por Hjelmslev no quinto e sexto capítulos acentua ainda mais o caminho em direção a uma epistemologia imanente. O teórico mesmo afirma que “nenhuma teoria científica pode ser elaborada sem uma colaboração ativa com a epistemologia” (2013, p. 17). Esse posicionamento contribui para que haja o estabelecimento da lógica implicativa que se instala aqui como uma diretiva do que virá a seguir: a limitação, apenas provisória, de seu objeto. Essa limitação leva em consideração a procura por uma constância que esteja fundamentada em aspectos extralinguísticos (p. 7), o que leva a uma exclusão provisória

do que é considerado variável. Tais variáveis retornariam, então, e “aqueles aspectos excluídos da primeira consideração são novamente introduzidos e assumem seu lugar num novo conjunto” (p. 24). Essa reintrodução dos elementos antes deixados de lado e seu reposicionamento em um novo conjunto pode ser interpretada como uma paráfrase do que virá no capítulo final: o estabelecimento de uma base nova e mais sólida ancorada sobre a imanência na qual “imanência e transcendência juntam-se numa unidade superior” (2013, p. 133). Aqui, ainda no sétimo capítulo, temos uma ocorrência das mais objetivas em relação à ascensão da teoria da linguagem a um nível epistemológico. Para organizar de maneira mais concisa nosso argumento, elaboramos a tabela abaixo que segue os mesmos princípios da anterior:

Tabela 3 – Ocorrências da imanência epistemológica nos *PTL*

Imanência epistemológica

“O desenvolvimento da linguagem está tão inextricavelmente ligado ao da personalidade de cada indivíduo, da terra natal, da nação, da humanidade, da própria vida, que é possível indagar-se se ela não passa de um simples reflexo ou se ela não é tudo isso: a própria fonte do desenvolvimento dessas coisas. ” (2013, p. 1, grifo nosso)	Cap. 1
---	--------

[...] uma constância que [...] se deixe projetar sobre a ‘realidade’ ambiente seja qual for a natureza desta [...] de modo que esta ‘realidade’ se ordene ao redor do centro de referência que é a linguagem [...] como um todo organizado que tem a estrutura linguística como princípio dominante. ” (2013, p. 7, grifo nosso)	Cap. 2
--	--------

“Isto na convicção de que nenhuma teoria científica pode ser elaborada sem uma colaboração ativa com a epistemologia ” (2013, p. 17, grifo nosso)	Cap. 5
---	--------

“A limitação (da teoria) pode ser considerada justificada se mais tarde permitir uma ampliação da perspectiva através de uma projeção da estrutura descoberta sobre os fenômenos que lhe são vizinhos, de modo tal que sejam explicados de maneira satisfatória à luz da própria estrutura [...]” (2013, p. 23, grifo nosso)	Cap. 7
--	--------

Fonte: elaborada pelo autor.

A análise dos nove capítulos iniciais nos permitiu chegar à conclusão de que os capítulos 3, 4, 6, 8 e 9 não possuem passagens substanciais da chamada imanência epistemológica e seus desenvolvimentos. Vejamos então pormenorizadamente as ocorrências que temos explicitados na tabela acima. A primeira ocorrência, ainda na primeira página do livro, serve de exemplo claro da ascensão de uma imanência

metodológica para uma epistemológica. Essa passagem é marcada pelo nivelamento inicial entre o desenvolvimento da linguagem e os fenômenos humanos, ambos em um mesmo nível, estatuto esse que é logo alterado: a própria língua seria a fonte de todas essas manifestações humanas. Esse é o primeiro apontamento para a virada epistemológica da imanência executada no último capítulo dos *PTL*. A segunda ocorrência é o detalhamento de como a língua formaria essa base/fonte: a partir da identificação de uma constância (sistema) que se deixe projetar sobre a “realidade”. Esse trecho pode ser tratado como uma paráfrase da formulação que encontramos no capítulo final: “Em vez de impedir a transcendência (realidade), a imanência (constância), pelo contrário, deu-lhe uma base nova e mais sólida” (Hjelmslev, 2013, p. 133). No capítulo cinco, Hjelmslev reforça a posição ativa da linguística – como teoria científica – frente a epistemologia e, por fim, a ampliação da perspectiva por meio de uma projeção da estrutura linguística pode ser considerada outra paráfrase da conclusão que Hjelmslev dá para a imanência no último capítulo. Logo, mesmo que tenhamos menos ocorrências da imanência epistemológica quando comparada com a metodológica, podemos identificar sua presença desde a primeira página da construção da teoria da linguagem nos *PTL*, dessa forma, a construção da primeira nunca deixa de lado também a construção da segunda.

Em vista disso, tais considerações formam a base do que acreditamos constituir o termo complexo, ou seja, a imanência enunciada nos *PTL* é formada – e modalizada – a partir da síntese de dois conceitos diferentes (cf. Lopes, 1997) no enunciado: imanência metodológica e imanência epistemológica. Dessa forma, temos, na enunciação pressuposta, um conceito complexo, ou a unidade de nível superior enunciada no capítulo 22 dos *PTL*. É importante retomarmos as definições conceituais que norteiam essa nomenclatura. É necessário apontar que a conciliação final – explícita – que Hjelmslev propõe para a imanência e para a transcendência revela uma nova perspectiva para a imanência:

A imanência e a transcendência juntam-se numa unidade superior baseada na imanência. A teoria linguística, por necessidade interna, é levada a reconhecer não apenas o sistema linguístico em seu esquema e seu uso, em sua totalidade assim como em seus detalhes, mas também através dela, a atingir o domínio do saber humano em sua totalidade (Hjelmslev, 2013, p. 132).

Tal unidade articula-se como uma hierarquia na qual podemos visualizar duas classes: o sistema linguístico (em seu esquema e seu uso) e o saber humano. À primeira

imanência, metodológica, caberia apenas a construção de uma teoria imanente articulada à primeira dessas classes. Enquanto à segunda, epistemológica, cabe a tarefa de articular dois planos: imanência metodológica (linguística) e transcendência (conhecimento humano). Aqui, cabe citar novamente a ambição gnosiológica da proposição hjelmsleviana citada por Badir (2014), ou seja, a teoria da linguagem, para Hjelmslev, deve ser capaz de fornecer sua própria classificação a partir do princípio da simplicidade, de modo que entendemos que uma das proposições elaboradas nos *PTL* é justamente essa que pretende alçar a teoria da linguagem a um nível epistemológico. Paralelamente, na obra *Sémantique structural* (1986), Greimas afirma que o mundo humano se forma a partir da significação, ou seja, só se define essencialmente a partir do crivo significativo humano (p. 11). Alinhamos, ainda, do mesmo modo, outra famosa afirmação do mestre lituano: “Fora do texto não há salvação: todo o texto, somente o texto, nada além do texto”³¹; ambas formulam uma base segundo a qual o mundo humano fenomenológico só é compreensível por meio da linguagem. Dentro dessa formulação do mundo, encontra-se a epistemologia das ciências gerais, esta que também só pode ser apreendida por meio da linguagem segundo o raciocínio de Greimas. Esse postulado fornece as bases para compreendê-la como a semântica científica, que, para Greimas, só se estabelece a partir de uma existência linguística dividida em três níveis – descritivo, metodológico e epistemológico (1986, p. 175).

Qualquer estudo que envolva a linguagem é, antes de tudo, metalinguístico, isto porque usa a língua para falar de si mesma. Quando a gramática normativa estabelece as classes gramaticais para o português brasileiro, estabelece-se um nível de análise, descritivo, que incorrerá sobre o nível analisado, objeto. Segundo Greimas (1986, p. 185), estabelece-se um nível para o objeto analisado – a língua – e outro para a metalinguagem descritiva que a analisa. Portanto, aos três níveis citados anteriormente – descritivo, metodológico e epistemológico – cabe adicionar um quarto: o da língua analisada, ou objeto. Se o nível do objeto é submetido a uma análise pelo nível descritivo, o nível metodológico regula os conceitos estabelecidos no nível descritivo, ou seja, também os analisa tornando-se assim uma meta-metalinguagem. No caso de uma gramática, essa relação se estabelece por meio das definições que regulam as classes gramaticais, por exemplo. Assim, uma definição de uma gramática do grego que estabeleça que diante de

³¹ N. T.: Traduzido do original em francês: “Hors du texte, point de salut. Tout le texte, rien que le texte et rien hors du texte” (Greimas, 1974, p. 25).

um substantivo, o artigo serve para designar de maneira precisa um ser, temos uma classe (artigo) do nível descritivo sendo analisada e precisada pelo nível metodológico. Para o último nível, deve-se estabelecer uma axiomática, ou seja, um nível regulador que estabeleça seus conceitos a partir de um número mínimo de elementos que sejam tão gerais quanto possível.

O primeiro encaminhamento, largamente intuitivo, consiste - partindo-se de um objeto assumido como cognoscível (a linguagem-objeto, em semiótica) - em elaborar, em primeiro lugar, uma linguagem de descrição e, a seguir, justificá-la com o auxílio de uma linguagem metodológica para encontrar, finalmente, o nível onde os conceitos, indefiníveis, e as hipóteses, não-demonstráveis, epistemológico deverão ser organizados em uma axiomática (Greimas, 2012, p. 456).

Dessarte, o mundo enquanto entidade somente apreensível por meio da linguagem, torna-se, ele próprio, uma linguagem-objeto submetida a uma análise metalinguística. Hjelmslev (2013) produz sua argumentação em prol de uma teoria da linguagem que não só define a si mesma, mas também é capaz de definir os conhecimentos do mundo, do mesmo modo que os níveis metalinguísticos estabelecidos por Greimas. Em conclusão, é necessário retomar a Tabela 3, na qual observamos que, nos *PTL*, esse nível terciário e metalinguístico é constituído, nos nove capítulos analisados, por uma metalinguagem que abrange, além do método linguístico – as definições da teoria da linguagem estabelecidas no livro –, também essa linguagem-objeto que é o mundo fenomenológico, incluindo aqui o próprio fazer epistemológico que é reasentado sobre uma base imanente e linguística. Assim, temos uma complexificação da imanência no sentido de que há duas proposições estabelecidas no texto: uma no nível metodológico, que regula a descrição, e outra no nível epistemológico, que busca reduzir de maneira arbitrária – mas adequada – as definições a um número mínimo de axiomas regulados pela junção da imanência e da transcendência em uma unidade superior baseada na imanência (Hjelmslev, 2013, p. 137). Acreditamos que esse postulado final, aliado a nossa análise, abre margem para que a imanência apresentada no último capítulo não seja vista como um ato repentino e brusco, mas sim como um desenvolvimento bem estabelecido na lógica discursiva do livro.

3.1.3 A imanência em A linguagem

A linguagem, quando analisada a partir da edição em inglês, divide-se em doze capítulos além da introdução do tradutor – Whitfield – e do índice remissivo. Assim como

agrupamos os capítulos dos *PTL* segundo as relações de conteúdo, por assim dizer, podemos indicar um procedimento semelhante para o livro em questão. Mantivemos a mesma organização de capítulos e partes em nossa tradução. Nesses termos, podemos propor a seguinte divisão. Os capítulos iniciais (1-5) - especialmente o capítulo das “Observações preliminares” - dedicam-se à apresentação de uma “linguística independente” (1970, p. 4) e assemelham-se, nesse sentido, ao primeiro capítulo dos *PTL*. Nesses capítulos, apresenta-se essencialmente o que já fora feito em linguística como um meio para obter outros conhecimentos e não como um fim em si mesma, imanente, elaborada a partir de um ponto de vista *linguístico* (p. 5), como é a proposta hjelmsleviana. Os capítulos seguintes (6-8) dão conta das relações entre línguas e a apresentação das famílias linguísticas, suas ramificações e parentescos. Por fim, os capítulos conclusivos (9-12) formulam os princípios da teoria da linguagem que embasa os procedimentos anteriores. Esse último conjunto é o que mais se aproxima do estilo formulado nos *PTL* e, por isso, o mais teórico também.

Com relação às ocorrências do vocábulo *imanência* e suas variantes, encontramos apenas uma em todo o livro, especificamente no capítulo “Relação tipológica das línguas”, o que ressalta o compromisso de Hjelmslev com uma apresentação da teoria mais operacional e menos calcada em definições abstratas. Além disso, como Whitfield ressalta ainda em seus comentários sobre a tradução, o termo *glossemática* é completamente apagado, não aparecendo sequer uma vez durante todo o livro. É possível localizar uma ocorrência de um termo que aparentemente deriva dele. No capítulo “Diferentes graus linguísticos”, vemos a apresentação do termo *glosses*³², que, à primeira vista, poderia sugerir a indicação de *glossemas*, porém, verifica-se, em comparação ao que encontramos na versão em inglês dos *PTL*, que em inglês esse termo apresenta-se como *glossemas* (Hjelmslev, 1969, p. 80), nunca como *glosses*. Verificamos, assim, que o uso dado a “glosa” em *A linguagem* se aproxima de seu uso corrente no vocabulário linguístico - uma tradução literal - e não de seu uso glossemático. Dessa forma, reitera-se a afirmação de Whitfield e dá-se um exemplo do que seria a nossa proposta para o comentário do presente capítulo: a partir da ausência do termo fundador – glossemática – constatada pelo tradutor, assim como de nossa constatação da ausência (apenas) aparente das formas mínimas que a teoria estabelece em sua base explicativa, comentaremos a construção da imanência em *A linguagem*. Para isso, procederemos a partir da

³² Traduzido como *glosas* em nossa versão.

apresentação de trechos em que acreditamos existir uma tendência para o apagamento de certas características mais específicas e definicionais da teoria da linguagem em favor de elementos mais gerais. Utilizaremos em nossa análise a versão em inglês cuja tradução estará disposta em notas de rodapé.

3.1.3.1 Escolhas lexicais

Em primeira instância, podemos perceber como Hjelmslev procura não entrar na discussão acerca das escolhas lexicais que faz ou mesmo dos desenvolvimentos que suas proposições podem trazer. Como exemplo disso, foram coletadas algumas passagens em que vemos o potencial para desenvolvimentos metateóricos que, todavia, não se observam. Assim: “A word is not borrowed merely because **the thing it designates** has no native designation. People borrow out of a desire to imitate, and borrowings are subject to all the caprices of fashion”³³ (Hjelmslev, 1970, p. 64, grifo nosso). Nota-se que há no trecho marcado o apelo para um suposto referente externo à língua para referir-se a algo interno ao sistema da língua – a palavra. É interessante notar como Hjelmslev não se preocupa em distanciar-se do chamado realismo ingênuo como faz nos *PTL* (2013, p. 28) ou da indução positivista (p. 13), nem, por outro lado, procura delegar a outrem possíveis problemáticas derivadas dessas escolhas terminológicas/lexicais como também o faz nos *PTL* (p. 11). Esse posicionamento reforça nossa proposição – derivada de Whitfield – de que a sistemática construída em *A linguagem* reflete, propositadamente, um condicionamento menor aos desenvolvimentos teóricos da glossemática, no sentido de que se preocupa menos com a terminologia específica e com uma organização mais formalizada, tomando como padrão uma linguagem mais simples e uma terminologia a qual chamaríamos de operacional, como foi discutido no item 3.1.1 sobre as funções.

Do mesmo modo, outro ponto em que podemos observar essa lógica é o que surge no capítulo “Parent languages”. Vejamos: “And, still hypothetically, this prehistoric society can be assigned some definite place or other on the basis of conjectures, say, about the **plants and animals that might have been designated by the words** in the language.”³⁴ (1970, p. 81, grifo nosso). Novamente não há preocupação em resolver uma

³³ “Uma palavra não é meramente tomada de empréstimo porque **designa alguma coisa** que não possui uma designação nativa. Os falantes tomam emprestadas palavras por um desejo de imitação e tais empréstimos estão sujeitos a toda sorte de caprichos de estilo”. Tradução nossa.

³⁴ “E, ainda hipoteticamente, a essa sociedade pré-histórica pode ser designado um determinado lugar ou outro com base em conjunturas, digamos, **baseadas nas plantas e animais que por ventura foram designados pelas palavras** nessa língua.” Tradução nossa.

possível interpretação de que Hjelmslev se afastaria do caráter *linguístico* preconizado por sua teoria ao citar elementos exteriores à língua. Esses casos trazem à tona novamente a questão da validação da teoria em relação aos dados da experiência. Como já salientamos nos capítulos anteriores, Hjelmslev afirma que a teoria em nada depende de tais dados para confirmar sua validade e sua existência como teoria. É somente com o texto que a teoria deve se ocupar por meio de uma descrição não contraditória e exaustiva (2013, p. 19).

Assim, apesar de não haver essa problematização em torno das proposições citadas assim que elas ocorrem, Hjelmslev define, sim, quais os rumos que a construção da teoria deve tomar e quais deve evitar. Quanto à base da teoria da estrutura linguística, o mestre dinamarquês nos diz que a experiência não pode estar no âmago da construção dessa teoria, tanto porque seria impossível investigar todos os textos possíveis como tal empreitada seria também fútil, visto que a teoria deve prover um cálculo geral para todas as possibilidades existente e futuras de textos (1970, p. 105). Além disso, a suposta designação de uma “coisa” terciária exterior à língua e pela qual é designada, segundo a perspectiva de Bevidas (2015), reflete não algo de fora, mas sim um referente internalizado à linguagem (p. 6), que não se apoia nem em uma perspectiva realista, nem transcendental, mas sim imanente.

Em síntese, podemos dizer que Hjelmslev não estaria permitindo a entrada de um objeto terciário e extralinguístico em sua teoria – apesar de não deixar claro isso – mas sim dando à *transcendência* uma nova base – mais sólida – assentada sobre a imanência, assim como o fez nos *PTL* (2013, p. 133), porém de uma maneira menos comprometida com as especificidades da teoria e mais voltada ao público leigo num primeiro contato com essa visada linguística. Essa perspectiva terciária, assentada sobre a imanência, que reinterpreta o que fora tratado como transcendente agora como parte da estrutura linguística é vista também em *A linguagem*. É um processo análogo e que acrescenta ao corpo da perspectiva que vê a ascensão da imanência a um patamar epistemológico. Podemos observar isso no seguinte trecho:

The content of language is nothing less than the world surrounding us, and the minimal particular meanings of a word, the particular meanings that are individuals (cf. p. 114) are the *things* of the world: the lamp that stands here on

my desk is a particular meaning of the word *lamp*; I myself am a particular meaning of the word *man*. (1970, p. 120)³⁵

Observa-se que o mundo e suas “coisas” se tornam parte do plano do conteúdo da língua, ou seja, fazem parte agora do que Hjelmslev chama de *metasemiótica*, ou uma semiótica cujo plano do conteúdo também é uma semiótica (2013, p. 121).³⁶ Citamos aqui os *PTL* pois em *A linguagem* o teórico não trata diretamente e também não nomeia essas categorias, mas podemos observar que elas fazem parte sim do seu pensamento, apenas são abordadas de uma maneira diferente, como ao definir que o plano do conteúdo de uma gramática é a expressão e o conteúdo de uma língua natural.³⁷ Ainda sobre isso, Hjelmslev afirma que não só o mundo que nos cerca é parte integrante da língua, mas também todas as ciências fazem parte dessa estrutura:

All sciences other than linguistics are actually theories of the linguistic content studied independently of the linguistic structure, just as physiological phonetics and physical acoustics are the study of the linguistic expression independently of the linguistic structure (1970, p. 120).

A partir de uma leitura comparativa, observa-se que é o mesmo movimento feito ao final do penúltimo capítulo dos *Prolegômenos* (Hjelmslev, 2013, p. 121). É dada a essa metasemiótica a função de analisar as múltiplas possibilidades do conteúdo – geográficos, históricos, políticos, sociais – que se ligam a aspectos nacionais, regionais etc. Todos esses elementos entram na estrutura linguística por meio da imanência, ou seja, fazem parte do plano do conteúdo de uma língua natural e como tal podem receber um tratamento e análise linguísticos.

Assim, é importante notar que o tratamento e a apresentação dados à glossemática em *A linguagem* não estão ausentes, mas formam, sim, uma estrutura subjacente aos desenvolvimentos do texto, que se apresenta por vezes de maneira menos formal – tomando aqui o sentido da palavra em termos hjelmslevianos – e, portanto, mais *operacional*.

³⁵ “O conteúdo da língua nada mais é que o mundo que nos cerca e os significados específicos e mínimos da palavra. Os significados particulares e individuais (veja a p. 120) são as *coisas* do mundo: a lâmpada que está aqui na minha mesa é um significado particular da palavra *lâmpada*; eu mesmo sou um significado particular da palavra *homem*” (tradução nossa).

³⁶ Cf. as definições sob o verbete “Mundo natural” do Dicionário de Semiótica (Greimas; Courtés, 2012, p. 291)

³⁷ Traduzido e parafraseado do trecho original: “What constitutes the content in grammar, the content plane of grammar, is the expression and content of the everyday language”. (Hjelmslev, 1970, p. 134)

3.1.3.2 Operacionalidade

Como ressaltamos anteriormente, as definições que surgem em *A linguagem*, em sua maioria, constituem o que Hjelmslev chama de definições *operacionais*, isto é, definições temporárias que podem ou não ser incorporadas à teoria – ou formalizadas – ou até mesmo eliminadas sem prejuízo para a construção do procedimento teórico (2013, p. 25). Discutimos sobre esse caráter provisório dessas definições anteriormente e, além de explorar algumas ocorrências de forma comparativa entre os *PTL* e *A linguagem*, aqui procederemos às ocorrências em que, no último, Hjelmslev aponta a operacionalidade do que é tratado. Pode-se encontrar, como veremos abaixo, cinco ocorrências – além daquela encontrada no índice remissivo – em que há referência direta às chamadas *definições operacionais*. Vejamos a tabela abaixo:

Tabela 4 – Definições operacionais em *A linguagem*

Original	Tradução
This is what is called an operational definition : it is a definition that provides a criterion, a test to decide whether genetic relationship is present in a given case or not (1970, p. 30, grifo nosso)	Isso é o que é conhecido como uma definição operacional : uma definição que provê um critério, um teste que decide quando uma relação genética se faz ou não presente em um determinado caso.
We have given an operational definition of genetic relationship above (p. 30) and shown that it consists in what we have called element-functions, constant correspondences between the systems of expression elements in the related languages (1970, p. 82, grifo nosso)	Algumas páginas acima, demos uma definição operacional da relação genética (p. 50) e demonstramos que ela consiste no que chamamos de elementos-funções, ou seja, correspondências constantes entre os sistemas dos elementos da expressão nas línguas relacionadas.
Now we have a solid ground of reality under our feet, and we can supplement our earlier operational definition of genetic relationship with a “ real ” definition : Genetic relationship between two languages means that they have developed from a single basic, or parent, language. (1970, p. 83, grifo nosso)	Agora temos uma base sólida e real embaixo de nossos pés e, assim, podemos suplementar nossa definição operacional anterior de relação genética com uma definição “real” : relações genéticas entre duas línguas significa dizer que ambas se desenvolveram a partir de uma mesma língua base/originária.

Having **operationally defined** genetic relationship as a function between languages that consists in element-functions, we may give a corresponding **operational definition** of typological relationship as a function between languages consisting in the fact that categories in each language have function to categories in each of the others. (1970, p. 95, grifo nosso)

Ao termos **definido operacionalmente** a relação genética como uma função entre línguas formada por elementos-função, podemos dar uma classificação correspondente para a **definição operacional** da relação tipológica como uma função entre línguas formada pelo fato de que categorias em cada língua têm uma função com categorias em cada uma das outras línguas.

Fonte: elaborada pelo autor.

A primeira ocorrência define um critério – ou testes – por meio do qual pode-se identificar ou não a presença de relações genéticas. Além disso, essa definição recebe um caráter *operacional*, ou seja, por ora possui um caráter temporário. A essa definição, das relações genéticas, Hjelmslev (1970, p. 30) dá o título de *continuação*. O teórico, no momento em que a define como operacional, dá a essa definição um caráter provisório, o que se verifica quando ele nos diz que a natureza dessa função ainda não foi minuciosamente inspecionada (1970, p. 30). Para entendermos melhor a extensão das definições elaboradas por Hjelmslev em sua teoria, é necessário analisar outras obras do teórico. Por exemplo, é interessante notar que não encontramos referência a essa definição na lista de definições dos *PTL* (2013). Ao final do livro, é organizada uma lista com 108 definições, na qual não se localiza menção à definição de *continuação*. Entretanto, encontramos-na no *Résumé d'une théorie du langage* (2015), texto conciso que fora concebido de modo a apresentar, sem exemplos práticos, as definições necessárias à formação da teoria da linguagem. Tais definições apresentam-se também sob a forma de um hipertexto em que há interligações entre as definições apresentadas, como veremos a seguir. Podemos identificar a definição apresentada em *A linguagem* como *continuation*, *changement sémiotique* (definição 207), além das definições 208, 209 e 210, respectivamente *laison génétique* (relação genética), *famille sémiotique* (família semiótica) e *diachronie* (diacronia) como exemplos de definições formais que retomam inteiramente a ideia abordada – *continuation; liaison génétique; famille sémiotique* – ou a expandem – *diachronie*.

Dessa forma, podemos supor que a ausência da definição para relações genéticas não apareça nos *PTL* em virtude do foco mais restrito e conciso quando

comparado à abordagem que encontramos em *A linguagem*. Enquanto este último concentra-se fundamentalmente sobre a relação genética que existe entre as línguas, além de ser organizado em torno dos desdobramentos possíveis que as diferentes relações podem acarretar; o primeiro, mais academicizado e formalizado, elabora uma esquematização geral dos fundamentos da glossemática. Diferentemente destes, o *Résumé* (2015) propõe-se a abarcar, na forma de uma descrição detalhada, todos os meandros da teoria. Mesmo que esse intento não tenha sido possível por completo, em vista dos problemas de saúde que acometeram Hjelmslev, como salienta Whitfield no prefácio da edição em inglês do *Résumé of a Theory of Language* (1975), pode-se verificar que a extensão deste recobre, e expande, o que fora desenvolvido tanto nos *PTL* como em *A linguagem*. Assim, podemos considerar que dentro da estrutura geral das obras de Hjelmslev, houve uma passagem de uma definição puramente operacional, a que encontramos para *continuação* em *A linguagem*, para uma formal, vista no *Résumé* (2015) sob a definição 207 como *continuation, changement sémiotique*, o que nos leva a concluir que essa definição, no livro mais amplo, faz parte definitivamente da organização da teoria.

Páginas à frente, agora no capítulo “Parent languages”, Hjelmslev retoma a definição que rege as relações genéticas – a *continuação* – e nos diz que essa categoria operacional é formada por funções-elemento que são correspondências constantes entre os sistemas da expressão das línguas relacionadas. Novamente, ao termo operacional não é dado nenhum contorno temporário. Todavia, na página seguinte, nossa terceira ocorrência, é demonstrado que a essa definição operacional deve sobrevir uma definição *real* (p. 83). Entendemos que esse “real” remeta à formalização da definição, agora assentada sobre uma base mais sólida. Esse real formalizado é constituído a partir das definições teóricas particulares da teoria da linguagem de Hjelmslev, que criam uma realidade particular que pode ser observada e descrita segundo os parâmetros estabelecidos pela teoria. A ideia de uma realidade objetiva apreensível em sua totalidade torna-se então ingênua e descabida, interessa, portanto, o constructo delimitado pela teoria em relação ao seu objeto (Dascal; Borges, 1991, p. 18). Dessa forma, podemos perceber o mesmo objetivo que fora demonstrado nos *PTL* – quando o teórico se debruça sobre o conceito de formalidade e operacionalidade das funções –, de forma concisa e intensa (2013, p. 25) alcançado aqui de forma mais profusa e extensa.

Por fim, no capítulo seguinte, “Typological relationship of languages”, temos as duas últimas ocorrências. A primeira retoma a *continuação* chamando-a novamente de

definição operacional. Apesar de termos visto a introdução de uma definição “real” em lugar dessa definição supostamente temporária, Hjelmslev volta a chamá-la pela terminologia menos formal. Uma vez que os *PTL* e *A linguagem* datam do mesmo período, sabemos que a dicotomia entre definições operacionais *versus* definições formais já estava bem definida, dada sua ocorrência nos *PTL*. Supomos então que a escolha por voltar ao termo “definição operacional” responda mais uma vez a questões enunciativas, *i. e.*, que vise o leitor mais iniciante construído no interior do texto. Ainda assim, vemos no par operacional/formal uma estratégia diferente dos outros elementos levantados, uma vez que não mantém o paralelismo com os *PTL* de uma evolução progressiva da formalização dos termos. É de se perguntar se isso reflete tão somente as escolhas enunciativas que mencionamos, ou se se pode sugerir uma progressão da reflexão glossemática entre *A linguagem* e os *PTL*.

3.1.3.3 Imanência

Dentro de nossa proposta, analisamos também como a imanência é apresentada dentro da teoria disposta em *A linguagem*. Como já dissemos acima, contamos com apenas uma ocorrência direta do termo e nenhuma do hiperônimo da teoria da linguagem – a glossemática. Entretanto podemos observar em outras passagens o que é nossa hipótese mais geral: a ausência (apenas) aparente das formas mínimas que a teoria estabelece em sua base explicativa. Constatamos nos capítulos anteriores a presença dessas formas mínimas que dão algum contorno para a teoria da linguagem e como essas formas se apresentam também nos *PTL*. Da imanência, temos que a única ocorrência explícita está localizada na seguinte passagem:

When we try to set up a linguistic typology from more penetrating and immanent points of view than have been chosen till now, we soon come to see that the entities which must be compared in the different languages are *categories*³⁸ (Hjelmslev, 1970, p. 95).

A imanência aqui não é introduzida de nenhuma maneira especial, mas sim como uma atitude contrária ao que era propagado até o momento nos estudos linguísticos.

³⁸ Quando tentamos organizar uma tipologia linguística a partir de pontos de vista mais profundos e imanentes do que os escolhidos até o momento, logo veremos que as entidades que devem ser comparadas nas diferentes línguas são *categorias*. Tradução nossa, grifos do autor.

Em relação a isso, Hjelmslev, ainda nas observações preliminares, disserta sobre o que motivava os estudos linguísticos até o momento: a língua como um meio de se obter conhecimentos fisiológicos, físicos, psicológicos, sociológicos, históricos etc.; nunca como um fim em si mesma, nunca como uma linguística *linguística*. Esse capítulo introdutório carrega semelhanças com o primeiro capítulo dos *PTL*, onde surge a primeira das dez ocorrências diretas da imanência no livro. Assim, essa linguística *linguística* é, de fato, a linguística imanente sob uma perspectiva menos formalizada, mas, ainda assim, imanente. Da mesma forma, outro trecho importante para essa compreensão é o seguinte:

Naturally, the analysis must not be a haphazard chopping up of the text but must be performed with an eye to the *relations* between its constituent parts (see p. 32). These relations may be of several kinds³⁹ (1970, p. 97).

A análise não deve ser somente uma decupagem do objeto em si, mas deve, por outro lado, levar em consideração as relações internas que constituem esse objeto. Essa perspectiva considera notadamente uma ambição *linguística* em que não se pode levar em consideração para a análise elementos exteriores, procedimento que se alinharia ao que o realismo ingênuo executaria (2013, p. 28). Esse realismo ocupar-se-ia de diversas decupagens possíveis, pois sua base é a experiência e não as relações – ou funções – contraídas pelo objeto. Dessa forma, finda-se por produzir um procedimento que não é exaustivo e chega-se a um problema de indução, pois uma única ocorrência fora do padrão colocaria em questão todo o procedimento, que teria que ser rearranjado. Em suma, é no procedimento contrário que o método *linguístico* – imanente – se apoia. É somente a partir da soma das dependências internas que o objeto analisado toma forma, e é a partir dessas dependências que a análise deve proceder: da classe para o componente até as mínimas consequências em que se respeite os princípios da não-contradição, da exaustividade e da simplicidade.

As definições, como demonstramos em relação à *continuação*, surgem no livro de maneira quase que descompromissada, sem muito alarde de seus desenvolvimentos perante a teoria. Um exemplo disso é a definição de *manifestação* que surge no livro em questão sem os contornos definicionais formalizados que a circundam nos *PTL*: “In such a case, where a presupposed entity is not **manifested** to us but can be

³⁹ “Naturalmente, a análise não deve ser um recorte ao acaso do texto, mas deve ser feita levando em conta as relações entre suas partes constituintes (veja a p. 51). Essas relações podem ser de diversos tipos”. Tradução nossa.

supplied we say that it is *latent*”⁴⁰ (Hjelmslev, 1970, p. 98, grifo nosso). A primeira aparição da definição, cuja natureza não é especificada – se operacional ou “real” – ocorre de maneira rápida e sem uma introdução que explique a importância do termo para a teoria. No livro mais acadêmico, Hjelmslev dá uma definição clara, apesar de provisória, para a *manifestação*:

Diremos, ainda, que o uso linguístico manifesta o esquema linguístico, e chamaremos de manifestação a função contraída pelo esquema e pelo uso. Provisoriamente, estes termos têm apenas um caráter operacional. (2013, p. 83)

Aqui, observamos basicamente dois movimentos: explicitar a definição e ao que ela se refere e, em segundo lugar, evidenciar o caráter operacional do termo, ou seja, que, por ora, não se está lidando com um termo formalizado na teoria. Não se observa o mesmo movimento em *A linguagem*, o que aclara nossa proposição de que, apesar de não serem diretamente manifestadas em sua complexidade – no sentido não teórico, no caso –, as formas mínimas estão de fato presentes no texto.

É importante, portanto, compreender que a construção da imanência dentro de *A linguagem* obedece a uma ordenação mais extensa, como nos referimos acima também ao analisarmos os conceitos de operacionalidade e formalidade, o que significa dizer que a apresentação das formas mínimas da glossemática surge de maneira menos concisa e, entretanto, também menos detalhada em relação às implicações que os termos têm dentro da teoria. Podemos observar que exemplos práticos se fazem mais presentes aqui do que nos *PTL*. Isso pode ser observado quando voltamos nossa atenção novamente ao que vem logo antes da definição de *manifestação*:

In the same way, if a man is called a husband, the existence of his wife is presupposed, even though we might never happen to meet her. In other words, we may be given only one entity together with an implicational relation, and on that basis we can *supply* the relation with another entity - not with any particular concrete entity that we could describe in detail, but with an entity about which we know only that it is something presupposed by the first⁴¹ (1970, p. 98).

⁴⁰ “Em tal caso, em que uma entidade pressuposta não está manifestada, mas pode ser suprida, diremos que é uma entidade latente”. Tradução nossa.

⁴¹ “Da mesma forma, se um homem é chamado de marido, a existência de sua esposa é pressuposta, mesmo que nunca venhamos a conhecê-la. Em outras palavras, podemos nos deparar com apenas uma entidade em uma relação implicativa e, baseando-nos nisso, poderemos *suprir* a relação não com qualquer outra entidade concreta em particular que poderíamos descrever em detalhe, mas com uma entidade sobre a qual sabemos apenas ser algo pressuposto pela primeira”. Tradução nossa, grifo do autor.

Além de servir de exemplo para a manifestação, o trecho acima revela também um paralelo com o conceito de *catálise* apresentado nos *PTL* quando Hjelmslev nos diz que “podemos suprir a relação com outra entidade”. Essa entidade ausente é suprida por meio da relação que é estabelecida entre a entidade presente e ela. No caso acima, a uma pessoa que receba o título de “marido”, pressupõe-se a existência de um cônjuge. Dessa forma, entre ambas as entidades existiria uma relação de interdependência, pois ambas pressupõem a existência da outra. Na prática da análise, a *catálise* é utilizada para acessar “certos furtivos inacessíveis ao conhecimento por outras vias” (Hjelmslev, 2013, p. 99). Nesse caso, não é qualquer categoria que pode ocupar o espaço que está por ora vazio, mas sim um “sincretismo irresolúvel de todas as grandezas que se poderia conceber para a ‘posição’ considerada na cadeia” (2013, p. 100). No caso do exemplo acima, podemos contar com o hiperônimo “cônjuge” para exemplificar esse sincretismo de possibilidades.

Assim, podemos concluir que, apesar da baixíssima ocorrência direta da imanência no nível do discurso, temos a presença constante dela desde a formulação da linguística *linguística* (1970, p. 5) – em muito semelhante aos apontamentos iniciais dos *PTL* – até a apresentação das definições de cunho operacional que explicitamos no capítulo anterior, bem como de exemplos concretos de aspectos da teoria que não chegam a ser explicitados em definição, como o caso da *catálise*. Percebe-se, portanto, uma tendência nas ocorrências analisadas de um deslizamento, como supomos inicialmente, do discurso para um nível que se adegue ao público inexperiente em matéria linguística. Isso é feito tanto por meio das escolhas lexicais, que muitas vezes não se comprometem tanto com as definições da teoria – ou não as especificam –, como é o caso da contração das funções, apresentado no capítulo homônimo; como também da escolha do próprio autor em não se aprofundar em desenvolvimentos ou alcances que suas escolhas, sejam lexicais ou teóricas, têm perante a teoria da linguagem como um todo.

CONCLUSÃO

À guisa de conclusão, faremos um breve percurso acerca do lugar de Hjelmslev no percurso atual da linguística e da semiótica. À luz da trajetória – do *continuum* – entre operacionalidade e formalização, é interessante questionarmo-nos qual o lugar de Hjelmslev, e de *A linguagem*, em uma época em que se pergunta para onde a linguística e a semiótica querem ir. Para tanto, é necessário compreendermos, a partir dos

apontamentos feitos neste trabalho, um Hjelmslev preocupado não só com a metateoria, mas também com a operacionalização dessa teoria. Desse modo, vemos que o trabalho empreendido em *A linguagem* possui um teor mais aplicado, no que diz respeito à apresentação de uma lógica de análise, com apresentação de exemplos e resultados. Podemos observar isso, como já apontamos anteriormente, já a partir da organização dos capítulos do livro.

Os capítulos iniciais (1-5) ressaltam uma “linguística independente” dos fatos considerados extralinguísticos (Hjelmslev, 1970, p. 4) e assemelham-se, portanto, ao primeiro capítulo dos *PTL*. Nesses capítulos, apresenta-se essencialmente que o estudo linguístico fora usado como um meio para se obter outros conhecimentos e não como um fim em si mesmo, imanente, elaborado a partir de um ponto de vista linguístico (1970, p. 5), como se propõe a perspectiva hjelmsleviana. Os capítulos seguintes (6-8) dão conta das relações entre línguas e a apresentação das famílias linguísticas, suas ramificações e parentescos, que são definidos a partir da formalização e contração funcional entre os *elementos da expressão*. Por meio dessas construções, é visto o procedimento de elaboração de fórmulas. Tais fórmulas podem ainda ser representadas por elementos escolhidos arbitrariamente e, assim, ao invés de produzir uma extensa formulação sobre as contrações e concordâncias entre um determinado elemento do gótico, do celta e do latim, pode-se optar por uma fórmula mais enxuta, como “i.e.*m” (1970, p. 16). Por fim, os capítulos restantes (9-12) formalizam os princípios da teoria da linguagem que embasam os procedimentos anteriores.

Dentro desse escopo, observamos que tal procedimento de resgate, retomada e reformulação é um movimento visto também em alguns periódicos que se destacam por esse trabalho. Um deles é o grupo de pesquisa *La Reconstruction*⁴², articulado pelo Institut Ferdinand de Saussure e que tem François Rastier como seu principal agente. O instituto é uma organização científica fundada em 1998 por Claude Lévi-Strauss, Jean Starobinski, Rudolf Engler e Tullio de Mauro. Nele, são desenvolvidos três programas: *Archives Ferdinand Saussure*, *Programme Sémantique des textes* e *Programme Sciences de la culture*. O *La Reconstruction* se alinha ao último e tem em seu projeto diversos interesses que se organizam a partir das diferentes acepções da palavra *reconstrução*: filológico, operacional, epistemológico, filosófico, ético e político. Todas as significações da palavra

⁴² Disponível em: <https://lareconstruction.fr/>.

trazem consigo o senso de retomar debates tidos por encerrados e, assim, trazer novos debates e perspectivas.

Em terras brasileiras, temos a *Revista de Estudos da Linguagem* que é mantida pela Faculdade de Letras e pelo Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais. Em seu último volume, ainda a ser publicado, trouxe como temática a linguística geral hoje, ou seja, trabalhos sobre uma linguística que se interessa por conhecer e esclarecer seus próprios fundamentos filosóficos, ou, em outras palavras, uma linguística de fundamentos.

Nos últimos anos, o interesse pela obra de Hjelmslev aumentou consideravelmente. Vemos isso a partir de cursos, palestras e trabalhos que retomam sua teoria. Podemos citar o livro publicado por Beividas, *Epistemologia discursiva* (2020), que trata sobre uma proposta terciária para a teoria do conhecimento humano baseada na linguística e que passa por Hjelmslev. Também temos recentemente o curso *Hjelmslev Hoje* (2018) organizado pelo Prof. Dr. José Roberto do Carmo Jr. e apresentado em quatro módulos no Grupo de Estudos Semióticos GES-USP. Ao primeiro e segundo módulos foi reservado o debate sobre os conceitos gerais e teóricos sobre a teoria hjelmsleviana, enquanto que para os dois últimos, reservou-se para o debate sobre perspectivas atuais e novas visadas. Neste último bloco, Carmo Jr. cita o conceito de enunciação que não é diretamente citado em Hjelmslev, mas pode ser recuperado de forma especulativa. Ainda, em trabalho mais recente e já citado anteriormente, temos a tese de Valério (2023) intitulada *A descrição glossemática de estruturas multissemióticas na chamada de capa*, que propõe a descrição semiológica das chamadas de capa em jornais impressos a partir dos conceitos teóricos e metodológicos de Hjelmslev.

Nesta seara, temos a *Estudos Semióticos* (EsSe), revista que possui publicações quadrimestrais no formato online pelo Programa de Pós-Graduação em Semiótica e Linguística Geral da FFLCH-USP. A revista veicula trabalhos da área de Semiótica e outras áreas aproximadas em português, francês, inglês, espanhol e italiano. A EsSe tem contribuído de forma contumaz com os estudos hjelmslevianos ou conexos, o que é percebido de forma mais clara quando analisamos as publicações de três números em específico: o terceiro número do volume 19 publicado em 2023 sobre a temática “A semiótica e suas fontes: um olhar metasemiótico”; o terceiro número do volume 16 publicado em 2020, “Semiótica e epistemologia” e o número especial da AJCS (Association des Jeunes Chercheurs en Sémiotique) do volume 13 publicado em 2017. O primeiro número é um dossiê temático voltado ao estudo das fontes da semiótica sob uma

perspectiva metasemiótica, ou seja, a partir de suas próprias conjecturas. A partir dessa revisão de bases, abre-se margem para que se revisitem os fundamentos epistemológicos das práticas semióticas (Cigana; Moreira; Sofia, 2023). O segundo número citado trabalha sobre os diálogos possíveis entre semiótica e as teorias do conhecimento. Por exemplo, analisam-se em um dos artigos as possibilidades de diálogo entre Hjelmslev e os pressupostos da filosofia da linguagem de Frege e Wittgenstein (Bonfim, 2020). O último número citado possui artigos que abordam desde uma revisão dos conceitos de arbitrariedade e de imanência até uma discussão sobre o lugar da categoria de repetição na semiótica hjelmsleviana (Lemos, 2017).

Cabe citar também números dedicados a autores que, direta ou indiretamente, são afluentes do pensamento hjelmsleviano: os dois números dedicados a Greimas em seu centenário: o número 2 do volume 13 publicado em 2013 e o número 1 do volume 14 publicado em 2018; o número 3 do volume 11 publicado em 2015, cujo conteúdo foi dedicado ao Groupe μ , grupo de linguistas e semioticistas belgas da Universidade de Liège; o volume 15 publicado em 2019 em tributo a Claude Zilberberg; o segundo número do volume 19 publicado no ano de 2023 como homenagem a Jean-Marie Floch e, ainda a ser publicado, o número 3 do volume 20, cuja temática versa sobre os quarenta anos da publicação de *Sobre o sentido II*, de Algirdas Julien Greimas.

Em resumo, é notável o crescimento e o interesse que se tem ganhado a teoria hjelmsleviana e nosso trabalho entra nesse âmbito de expansão e reformulação que merece cada vez mais atenção dos teóricos em linguística e semiótica. Nesse contexto, é importante que seja ressaltada a atualidade do teórico para o debate acerca dos problemas em linguística geral e seus princípios. O grande nível de formalização que perpassa as obras de Hjelmslev acaba as colocando em um plano demasiado epistemológico e distante de uma lógica aplicacional direta, o que acarretou em diversas críticas, como salienta Jovanović (1991). Segundo ele, a glossemática fora acusada, “entre outras coisas, de considerar que a relação entre a teoria e o seu objeto tem caráter epistemológico-metodológico que não leva em consideração a funcionalidade da teoria” (Jovanovic, 1991, p. 183). Numa perspectiva histórica, pode-se considerar que a tais discussões epistemológicas será dada, em um primeiro momento, uma consistência metodológica a partir de Greimas (1973 [1966]) e a semiótica. Porém, é justo ressaltar a importância de *A linguagem* nesse quadro, por demonstrar que, já em 1943, houve uma elaboração que se preocupava em elaborar uma imanência epistemológica e, ao mesmo tempo, aplicar essa imanência a dados. Assim, nosso trabalho traz para a cena linguística e semiótica

uma nova perspectiva que vê um Hjelmslev que vai além das concepções puramente teóricas e metateóricas de sua glossemática e se debruça também sobre os problemas ditos concretos. Dessarte, estudar Hjelmslev hoje não é dar um passo atrás ou, ainda, lidar com abstrações sem aplicabilidade, mas trazer novas possibilidades para debates vistos como encerrados e, assim, abrir margem para novos desenvolvimentos.

À luz desse breve percurso que delineamos aqui, encontra-se nossa proposta de tradução que se propôs, inicialmente, como uma contribuição aos estudos linguísticos, em especial aos estudos hjelmslevianos, que careciam de uma tradução em nossa língua dessa obra do mestre dinamarquês, uma vez que não temos um acervo amplo de obras e trabalhos de Louis Trolle Hjelmslev publicados em português. Nesse ponto assentamos a legitimidade de nosso trabalho. Com a tradução, outros desenvolvimentos foram necessários: a contribuição das outras edições em línguas latinas, a expansão das fórmulas para incluir o português brasileiro e o acréscimo de exemplos – ou a substituição destes.

Além da contribuição à qual nos propusemos com a tradução, acrescentamos o estudo comparativo entre duas das principais obras do autor: *A linguagem* – em nossa tradução – e os *Prolegômenos a uma teoria da linguagem*. Dessa comparação pode-se lançar uma nova perspectiva para a construção dos dois livros e que pode, e deve, ser expandida e analisada em estudos posteriores. Nosso intento foi o de abrir um caminho inicial para que outros venham e possam, agora em nossa língua, construir novos saberes.

O profundo trabalho implicado na tradução da obra nos abriu as portas para a observação cuidadosa da construção discursiva da teoria da linguagem que subjaz aos dados e temas abordados em *A linguagem*. Ao discutir escolhas lexicais, a presença de definições operatórias (frequentemente no lugar de definições formais), a presença não declarada de operações da glossemática (como a catálise), além da própria inserção do termo relacionado à imanência, pudemos observar que, mesmo em um trabalho que visa a difusão a um público inexperto, Hjelmslev não deixa de aderir às bases da teoria da linguagem que veio delineando ao longo de sua trajetória. O que se verifica é, sim, uma série de estratégias de apagamento dos aspectos mais formalizados da teoria, o que justifica a interpretação de que essa obra visa um tal público mais iniciante.

Nossa análise de *A linguagem* encontra, dessa forma, uma construção menos concisa e, por isso, menos teorizada – quando comparada aos *PTL* –, o que se confirma nas ocorrências de termos e trechos escolhidas e analisadas, com construções mais extensas e átonas, se comparadas aos *PTL*, ainda que, como argumentamos, os próprios

PTL desenvolvam internamente sua própria dimensão implicativa na direção da construção necessária de uma imanência epistemológica.

A glossemática nas entrelinhas, formulação expressa em nosso título e trabalhada ao longo do cap. 3, surge então, em um primeiro vislumbre, como essa construção menos aparente, mas sempre presente, da constância teórica que perpassa como um fio condutor toda a obra de Hjelmslev, não como um objeto discreto, mas sim em um *continuum* estabelecido entre operacionalidade e formalização. Assim, mesmo quando temos uma construção discursiva em que a glossemática e sua organização teórica é menos aparente, como é o caso de *A linguagem*, é possível depreender ainda a imanência, ponto central da glossemática, e seus elementos constitutivos. Em uma segunda análise, essa glossemática sutil, entre as linhas, pode ser entendida também como a construção elaborada da tradução de *A linguagem* entre as diferentes línguas e as escolhas que foram feitas em relação a apresentação de seus elementos teóricos. É uma questão interessante que pode ser desenvolvida em estudos posteriores.

Naturalmente, não nos propusemos a um estudo exaustivo dessas ocorrências, dada a extensão do gênero em que executamos nossa análise. A expansão dessa análise poderá vir em trabalhos de maior fôlego que se proponham a esse intento. Assim, salientamos que a compreensão da extensão do pensamento hjelmsleviano é um trabalho sempre em devir: em constante transformação, evolução e movimento. Nosso intuito foi o de abrir alguma margem para que esse trabalho e esse movimento ganhem ainda mais alcance e escopo a partir dos estudiosos em linguística de nosso país. Com a tradução de *A linguagem*, esperamos que outros trabalhos venham para dar corpo e conteúdo aos estudos linguísticos gerais e semióticos em nossa língua.

REFERÊNCIAS

- ALTMAN, Cristina. **A guerra fria estruturalista**. São Paulo: Parábola, 2021.
- BADIR, Sémir. Inmanencia y empirismo. Examen epistemológico de la teoría del lenguaje de Louis Hjelmslev. **Tópicos del seminario**, v. 31, 2014, p. 71-93.
- BARROS, A. F. de, *et al.* A Realização das Vibrantes do “Globês” no Português Brasileiro. **Revista de Letras e Artes**, v. 17, n. 1, jan./ jun. 2017.
- BEIVIDAS, Waldir. A teoria da linguagem de Hjelmslev: uma epistemologia imanente do conhecimento. **Estudos Semióticos**, v. 11, n. 1, 2015, p. 1-10.
- BEIVIDAS, Waldir. **Epistemologia discursiva: a semiologia de Saussure e a semiótica de Greimas como terceira via do conhecimento**. São Paulo: FFLCH/USP, 2020.
- BERMAN, Antoine. **A Tradução e a Letra ou o Albergue do Longínquo**. Florianópolis: PGET/UFSC, 2013.
- BONFIM, Jonatan Henrique Pinho. Pressupostos filosóficos da teoria da linguagem de Hjelmslev: a Filosofia Analítica de Frege e Wittgenstein. **Estudos Semióticos**, v. 16, n. 3, p. 1-17, 2020. Disponível em: <https://www.revistas.usp.br/esse/article/view/172828/165115>. Acesso em: 01 dez. 2024.
- CÂMARA JR., J. M. **Dicionário de Linguística e Gramática**. Petrópolis: Vozes, 2011.
- CARMO JUNIOR, José Roberto do. **Curso Hjelmslev Hoje 3**. Youtube, 27 de dez. de 2018, 1h52. Disponível em <https://www.youtube.com/watch?v=Mx5cYwZNRuE>. Acesso em: 18 de mai. 2024.
- CARMO JUNIOR, José Roberto do. Estratégias enunciativas na produção do texto publicitário verbo-visual. *In*: OLIVEIRA, Ana Cláudia; TEIXEIRA, Lucia (org.). **Linguagens na comunicação: desenvolvimentos de semiótica sincrética**. São Paulo: Estação das Letras e Cores, 2009. p. 169-184.
- CARMO JR., José Roberto do. **Melodia & prosódia: um modelo para a interface música-fala com base no estudo comparado do aparelho fonador e dos instrumentos musicais reais e virtuais**. 2007. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8139/tde-12112007-141109/pt-br.php>. Acesso em: 21 out. 2023.
- CHOMSKY, Noam. **Syntactic Structures**. New York: Mouton de Gruyter, 2002.
- CIGANA, Lorenzo; MOREIRA, Patricia; SOFIA, Estanislao. A semiótica e suas fontes: um olhar metasemiótico. **Estudos Semióticos**, v. 19, n.3, 2023. Disponível em: <https://www.revistas.usp.br/esse/article/view/220029/201474>. Acesso em: 01 dez. 2024.

CIGANA, Lorenzo. La notion de “participation” chez Louis Hjelmslev: un fil rouge de la glossématique. *In: Cahiers Ferdinand de Saussure*, n. 67, p. 191-202, 2014.

DASCAL, Marcelo; BORGES NETO, José. De que trata a linguística, afinal?. *In: Histoire Épistémologie Langage*, v. 13, n. 1, 1991, pp. 13-50.

FARACO, Carlos Alberto. Estudos pré-saussurianos. *In: MUSSALIN, Fernanda ; BENTES, Anna Christina. Introdução à linguística: fundamentos epistemológicos*, vol. 3. São Paulo: Cortez, 2004, p. 27 - 52.

FAÇANHA, Vinicius. **Condições semióticas da totalidade** : a dupla leitura de Os dragões não conhecem o paraíso. 2024. Dissertação (Mestrado em Linguística) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Linguística, Fortaleza, 2024. Disponível em:
https://repositorio.ufc.br/bitstream/riufc/75922/3/2023_dis_vfcsousa.pdf

FIORIN, José Luiz. Semiótica e história. *In: Seminários Teóricos Interdisciplinares do SEMIOTEC*. Belo Horizonte, v.1, n.1, p. 1-31, nov. 2012. Disponível em:
www.periodicos.letras.ufmg.br/index.php/stis/issue/view/120. Acesso em: 21 out. 2023.

FONTANILLE, Jacques; GREIMAS, Algirdas Julien. **Semiótica das paixões**: dos estados de coisas aos estados de alma. Trad. Maria José Rodrigues Coracini. São Paulo: Ática, 1993a.

FONTANILLE, Jacques; ZILBERBERG, Claude. **Tensão e significação**. Trad. Ivã Carlos Lopes, Luiz Tatit e Waldir Bevidas São Paulo: Humanitas, 2001.

GREENBERG, Joseph H. **Indo-European and Its Closest Relatives**: The Eurasiatic Language Family, v. 1 Grammar. Redwood City: Stanford University Press, 2000.

GREIMAS, Algirdas J.; COURTÉS. **Dicionário de semiótica**. São Paulo: Cultrix, 2012.

GREIMAS, Algirdas J. L'Énonciation: une posture épistémologique. *In: Significação: Revista Brasileira de Semiótica*, p. 9-25, 1974.

GREIMAS, Algirdas J. **Semântica estrutural**. Trad. Haquira Osakape e Izidoro Blikstein. São Paulo: Cultrix, 1973.

GUIMARÃES, Maximiliano. **Os fundamentos da teoria linguística de Chomsky**. Petrópolis: Vozes, 2017.

HJELMSLEV, Louis Trolle. Animé et inanimé, personnel et non-personnel e Sémantique structurale. *In: Essais linguistiques I: Travaux du Cercle Linguistique de Copenhague Nordisk Sprog- og Kulturforlag*, Copenhague, 1956, p. 211–250.

HJELMSLEV, Louis Trolle. Compte-rendu de Foundations of Language, de Louis Gray, *Acta Linguistica*, v. 2, n. 2, 1940, p. 122-126. Disponível em:
<http://dx.doi.org/10.1080/03740463.1941.10415374>. Acesso em: 27 abr. 2022.

HJELMSLEV, Louis Trolle. Edwad Sapir, **Acta linguistica**, v. 1, n. 1, 1939, p. 76-77. Disponível em: <http://dx.doi.org/10.1080/03740463.1939.10410849>. Consultado em: 29 abr. 2022.

HJELMSLEV, Louis Trolle. Essai d'une théorie des morphèmes. *In: Essais linguistiques I: Travaux du Cercle linguistique de Copenhague*. Nordisk Sprog- og Kulturforlag, Copenhagen, 1959, p. 152–164.

HJELMSLEV, Louis Trolle. **El lenguaje**. Trad. María Victoria Catalina. Madri: Gredos, 1968b.

HJELMSLEV, Louis Trolle. **Il linguaggio**. Trad. Anna Debenedetti Woolf. Torino: Einaudi, 1970.

HJELMSLEV, Louis Trolle. **La catégorie des cas: étude de grammaire générale**. Aarhus: Aarhus Universitetsforlag, 1935.

HJELMSLEV, Louis Trolle. La conception linguistique moderne. *In: Cahiers Ferdinand de Saussure*, n. 68, p. 223-226, 2015.

HJELMSLEV, Louis Trolle. **Language**. Trad. Francis j. Whitfield. Madison: The University of Wisconsin Press, 1970.

HJELMSLEV, Louis Trolle. La structure des oppositions dans la langue. **Onzième Congrès international de psychologie**, 1937, p. 241-242.

HJELMSLEV, Louis Trolle. **Le langage**. Trad. Michel Olsen. Paris: Les Éditions de Minuit, 1966.

HJELMSLEV, Louis Trolle. Omkring Sprogteoriens Grundlæggelse. *In: Travaux du Cercle Linguistique de Copenhague*, Nordisk Sprog- og Kulturforlag, Copenhagen, vol. 25, 1993b.

HJELMSLEV, Louis Trolle. Outline of the Danish expression system with special reference to the stød. *In: Essais linguistiques II: Travaux du Cercle Linguistique de Copenhague* Nordisk Sprog- og Kulturforlag, Copenhagen, 1973, p. 247–266.

HJELMSLEV, Louis Trolle. **Principes de grammaire Générale**. København: Danske Videnskabernes Selskab, 1968a.

HJELMSLEV, Louis Trolle. **Prolegômenos a uma teoria da linguagem**. Trad. J. Teixeira Coelho Neto. São Paulo: Perspectiva, 2013.

HJELMSLEV, Louis Trolle. **Prolegomena to a theory of language**. Trad. Francis J. Whitfield. Madison: The University of Wisconsin Press, 1969.

HJELMSLEV, Louis Trolle. **Résumé d'une théorie du langage**. Tradução e edição eletrônica de Alain Herreman, 2014.

- HJELMSLEV, Louis Trolle. **Sproget**: en introduktion. København: Berlingske leksikon, 1973.
- HJELMSLEV, Louis Trolle. Structure générale des corrélations linguistiques. In: **Essais linguistiques II**: Travaux du Cercle linguistique de Copenhague, Nordisk Sprog- og Kulturforlag, Copenhague, 1973, p. 57–98.
- JOVANOVIĆ, Aleksandar. Hjelmlev: um nome fundamental para a linguística do século XX. In: **Revista USP**. São Paulo: Faculdade de Educação, Universidade de São Paulo, n. 11, p.183-185, set./nov. 1991. Disponível em: <https://repositorio.usp.br/item/000869155>. Acesso em 19 mai. 2024.
- LEMOS, Carolina Lindenberg. **Condições semióticas da repetição**. 2015. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8139/tde-09062015-111352/pt-br.php>. Acesso em: 21 out. 2023.
- LEMOS, Carolina Lindenberg. La place de la répétition dans la sémiotique hjelmlevienne. In: **Revista de Estudos Semióticos**, v.13(spe), 2017, p. 18-29. Disponível em: <https://www.revistas.usp.br/esse/article/view/140733>. Acesso em: 21 out. 2023.
- LEMOS, Carolina Lindenberg; MAFRA, Matheus Henrique. A canção na trama do disco: relações de classe e componente. In: **Cadernos de Estudos Linguísticos**, Campinas, SP, v.65, n.00, 2023, p. 1- 15. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8670378>. Acesso em: 21 out. 2023.
- LOPES, Edward. **Identidade e diferença**: raízes históricas das teorias estruturais da narrativa. São Paulo: Editora da Universidade de São Paulo, 1997.
- MARTINET, André. Au sujet des Fondements de la théorie linguistique de Louis Hjelmlev. **Bulletin de la Société de Linguistique de Paris**, v. 42, 1946, p. 19-42. Disponível em: <https://gallica.bnf.fr/ark:/12148/cb343492767/date19460101#resultat-id-1>. Acesso em: 13 ago. 2022.
- PEIXOTO DA FONSECA, F. V. Aspectos do vocabulário e semântica do português arcaico. *Linguística*, 29(1), 1989, p. 129–136. <https://doi.org/10.4312/linguistica.29.1.129-136>. Consultado em: 20 jul. 2023.
- POPPER, Karl. **A lógica da pesquisa científica**. Trad. Leonidas Hegenberg e Octanny Silveira da Mota. São Paulo: Cultrix, 2008.
- PRADO, Maria Goreti Silva. O corpo próprio na semiótica. **Estudos Semióticos**, São Paulo (SP), v. 9, nº 1, p. 68-79, 2013.
- RAGON, Eloi. **Gramática grega**. Tradução de Cecília Bartalotti. São Paulo: Odysseus Editora, 2011.

RASTIER, François. Saussure e o desenvolvimento da semiótica. [Entrevista concedida a] ASEL. **Acta semiótica et lingvistica**, João Pessoa, vol. 22, 24 de out. de 2017.

SAPIR, Edward. **A linguagem**: introdução ao estudo da fala. Tradução de Joaquim Mattoso Camara Jr. Rio de Janeiro: Instituto Nacional do Livro, 1954.

SAUSSURE, Ferdinand de. **Curso de linguística geral**. Trad. de Antônio Chelini, José Paulo Paes e Izidoro Blikstein. São Paulo: Cultrix, 2012.

VALÉRIO, Yvanowik Dantas. **A descrição glossemática de estruturas multissemióticas na chamada de capa**. 2023. Tese (Doutorado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2023. Disponível em: <https://repositorio.ufc.br/handle/riufc/71539>. Acesso em: 21 out. 2023.

WHAT are the largest language families? **Ethnologue**, 2019. Disponível em: ethnologue.com/guides/largest-families. Acesso em: 26 jun. 2022.

WHITFIELD, Francis J. Louis Hjelmslev. **Language**, v. 42, n. 3, parte 1, 1966, p. 615-619. Disponível em: <https://gallica.bnf.fr/ark:/12148/cb343492767/date19460101#resultat-id-1>. Acesso em: 13 ago. 2022.

ZILBERBERG, Claude. **Elementos de semiótica tensiva**. Trad. Ivã Carlos Lopes, Luiz Tatit, Waldir Bevidas. São Paulo: Ateliê Editorial, 2011.

ZINNA, Alessandro. La inmanencia: línea de fuga semiótica. **Tópicos del Seminario**, Benemérita Universidad Autónoma de Puebla, Puebla, n. 31, jan. – jun., 2014, pp. 19-47.

APÊNDICE A - ORIGINAL - *Language: An Introduction****Contents***

Translator's Introduction
Prefatory Remarks
Linguistic Function
Genetic Relationship of Languages
Linguistic Structure and Linguistic Usage
Sign Formation
Language Families
Parent Languages
Typological Relationship of Languages
Types of Linguistic Structure
Types of Linguistic Usage
Linguistic Change
Languages of Different Degrees
Index

Translator's Introduction

Although the Danish original of this book (entitled *Sproget: En introduktion*) was not published until December 3, 1963, it was actually composed some twenty years earlier. It thus belongs to that extraordinarily creative period in Hjelmslev's life – resplendent with his vision of a new, “*linguistic linguistics*” – that also produced the *Prolegomena to a Theory of Language (Omkring sprogteoriens grundlæggelse)* and, among works now awaiting posthumous publication, the *Résumé of a Theory of Language (Sprogteori: Résumé)*. In contrast to the *Résumé's*, compendious, formal exposition of the “glossematic” theory, both the *Prolegomena* and *Language* were conceived as popularizations: the former as an outline, addressed to a scholarly audience, of the foundations of *glossematics*; the latter (in which the word *glossematics* does not once appear) as an introduction, primarily for laymen, into the problems of language and its scientific study.

An immediately striking contrast between the two “popular” works is to be accounted for by this difference of intended audience. A large portion of *Language* is concerned with the subject of genetic relationship between languages, and the structure of the whole book is determined by the central importance attached to the different possible relationships, genetic and nongenetic, that languages may have with one another. In the *Prolegomena*, on the other hand, attention is mainly directed to relationships within the individual language and within its texts. If, however, the reader is tempted to view the two books as in some way dividing the fields of comparative and descriptive linguistics between them, he courts serious danger of misreading them both. For one thing, the several discussions of internal relationships that he will find in the present work must be seen as necessary, organic parts of its argument. For another, it would be a superficial reading (however common) of the *Prolegomena* that ignored the final chapter and its express warning that “it is not the individual language alone that is the object of the linguist.”

To Hjelmslev it was obvious that linguistics must be “comparative,” not for mere comparison's sake, but for the successful pursuit of its true aims. The conclusion of the *Prolegomena* makes this quite apparent, even though the body of that work is devoted – as it could be devoted, in view of its special audience – to other aspects of the linguistic science that the author envisioned. It was equally obvious to Hjelmslev that an appropriate introduction to the study of language, both for the layman and for the linguist who wishes

to review the bases of his discipline, must be through the portal of comparative grammar – in particular, the comparative grammar of Indo-European. In his review⁴³ of Louis H. Gray's *Foundations of Language* he had written unequivocally:

Un mérite qui sera accueilli sans doute avec une satisfaction particulière est qu'il est tenu compte dans une très large mesure des faits et des théories indo-européens. L'indo-européen reste le stock où le linguiste de tous temps puisera de préférence ses matériaux et ses exemples; en outre l'indo-européen est le seul domaine linguistique que nous *croiyons* [italics added – FJW] vraiment connaître à fond. Il est donc oiseux de vouloir éviter au lecteur les difficultés techniques de la grammaire comparée; ce serait le tromper et lui dérober l'essentiel; il faut au contraire saisir le taureau par les cornes et faire ce que M. Gray a fait: introduire dans les problèmes de la linguistique générale en les regardant à travers les faits indo-européens. Ce procédé seul permettra de faire comprendre au lecteur que malgré tout la linguistique est une science, sinon dans la théorie, du moins dans la technique, et de le laisser pénétrer, à titre d'observateur, dans le laboratoire du savant. Rien de plus dangereux que de dissimuler au spectateur les difficultés réelles, et de lui faire croire qu'on peut étudier avec profit la linguistique générale sans des connaissances de grammaire comparée. Les conséquences de ces mirages ne laissent pas de se faire sentir dans le monde scientifique même, et contribuent largement à aggraver le schisme entre deux disciplines qui ont constamment besoin l'une de l'autre, et à répandre le dilettantisme dans la linguistique théorique. Linguistique Générale sans grammaire comparée est une linguistique em l'air.

The review of Gray must be read in its entirety by anyone who would learn more of Hjelmslev's preparation for the task he was to undertake in writing *Language*. At the very beginning of the review, he enumerates, without any claim to having made a complete list, twenty-six general works on language that had appeared in the nineteenth century, and thirty-one in the first four decades of the twentieth, and he acknowledges the need to which this accelerating production is a response. But the difficulties of meeting that need are enormous:

Les fondements de la linguistique sont controversés, aujourd'hui plus que jamais. Les exploits théoriques n'égalent pas le besoin pratique d'un traité d'ensemble embrassant toutes les branches de notre science d'un point de vue unique et communément adopté. La linguistique peut devenir une science; elle ne l'est pas encore; les qualités essentielles d'une science, l'objectivité et la précision, lui manquent encore dans une très large mesure. Les exposés dont on vient de parler en portent solvante fatalement l'empreinte; beaucoup d'entre

⁴³ *Acta Linguistica* 2.1 (1940-41): 122-26.

eux souffrent de subjectivité et d'exclusivité; souvent ils ont été écrits pour propager des vues particulières, bien que souvent originales; il en est qui sont des guides franchement infidèles... D'entre les nombreux exposés qui existent il n'y a donc guère un seul qui puisse servir comme véritable Manuel au non-initié ou au débutant, ni comme guide principal au lecteur averti.

Even Gray, despite the incontestable merits of his book, is found to have been led by his impartial attitude into a certain eclecticism “qui ne peut pas toujours réussir,” and the reviewer concludes:

Publier un livre sur la linguistique est un travail vraiment ingrat, et c'est la linguistique et non l'auteur qui est en faute. A moins de vouloir viser le but suprême qui consiste à présenter une doctrine purement personnelle et en même temps solide (et à notre avis il n'y a que F. de Saussure qui y ait réussi complètement) on est réduit à présenter un compromis où tout ne peut pas être au même niveau.

There can be no doubt that when Hjelmslev himself decided to undertake the “thankless task,” even though within the miniature frames of what he called a popular survey, it was the “supreme goal” that he aimed at and it was the model of Saussure that he kept in view. *Language* is a confessedly personal book that nevertheless boldly asserts its claim to being a solid one, free of the subjectivity for which Hjelmslev had reproached others.

The reader may find a paradox here, if not a contradiction, but Hjelmslev sets forth the claim with confidence, relying, above all, on his care in distinguishing between “what we know with certainty and what we do not know with certainty” (*hvad vi med sikkerhed ved og hvad vi ikke med sikkerhed ved*). This phrase, interestingly enough, finds its echo in the centennial lecture that Hjelmslev delivered in honor of his eminent predecessor Vilhelm Thomsen,⁴⁴ where we read: “He wished to distinguish, and distinguished everywhere as sharply as possible, between, as he says, ‘what can be known’ and ‘what must remain only rough hypothesis’: between what one knows with certainty and what one does not know with certainty.” The same lecture also gives strong indication of another respect in which Thomsen served as an exemplar, and that is in the matter of style. In discussing Thomsen’s *History of Linguistic Science*, Hjelmslev observes: “The style is unpretentious, plain, and simple, but by its very artlessness stands forth with an extra ordinary dignity and a quiet distinction. The elegance lies in the content alone, never

⁴⁴ “Vilhelm Thomsen. Foredrag paa Københavns universitet 26. Januar 1942,” *Gads danske Magasin* (1942), pp. 136-47.

in the form; the form is the simplest possible, and its merit lies precisely in the fact that it does not assert itself, but imperceptibly, faultlessly, smoothly, provides just what is needed for the content to act on the listener with all its carefully calculated weight. A mild and friendly tone pervades the exposition, a subdued humor lies ready to give color where occasion naturally arises.” A better characterization of Hjelmslev’s own style in *Sproget* I could not imagine. Even the briefest discussion of the background of this book would be incomplete without special mention of one more linguist: Edward Sapir. In Hjelmslev’s necrology of Sapir,⁴⁵ it is again the scholar’s respect for facts—together with the use that he makes of them—that is emphasized:

Sapir was constantly on his guard against hasty generalizations and sentimental prejudices; but this attitude did not involve sterile inductivism; Sapir was a highly visionary and synthetic mind, but his general theories were constantly checked by his knowledge of facts. Sapir’s method was empiricism in the best sense of the word: his doctrine was theory built on experience. The reader of Sapir’s works has the constant feeling that facts are not stated in order to provide him with knowledge for its own sake, but only to provide him with the material necessary for thinking... Sapir’s theory of language inspires with confidence because the only confidence he has in it himself is in the facts. For this reason, Sapir’s *Language* (1921) is likely to hold its place as a classical work and as one of the very best introductions to general linguistics yet written. The present writer clearly remembers the time when he first read it; it was a revelation, a confirmation of his vague anticipation of the possibility of establishing a *comparative general linguistics* destined to supersede the subjective and sentimental *philosophy of language* of the past.

Not surprisingly, Sapir’s *Language* heads the list of suggested reading appended to Hjelmslev’s own book. Although, out of considerations of appropriateness, I have elected not to reproduce this list in the English edition, it is interesting to examine the titles that it includes of works published before 1944. Among the recommended general introductions, Sapir’s is followed by Vendryes’ *Le langage*, Sandfeld’s *Sprogvidenskaben*, Marouzeau’s *La linguistique*, Jespersen’s *Philosophy of Grammar*, Pedersen’s *Sprogvidenskaben i det nittende århundrede*, Graff’s *Language and Languages*, Bloomfield’s *Language*, Gray’s *Foundations of Language*, Skautrup’s “Sproget” (in *Alverdens Viden III*) and Collinder’s *Introduktion i sprogvidenskaben*. Saussure’s *Cours de linguistique générale* appears, curiously enough, not among the

⁴⁵ *Acta Linguistica* 1.1 (1939): 76-77.

general works, but in the references under “Linguistic Change,” where special mention is also made of his “Mémoire sur le système primitive des voyelles dans les langues indo-européennes.” The only reference under “Sign Formation” is to Grammont’s *Traité de phonétique*, and the only under “Linguistic Structure and Linguistic Usage” is to Gardiner’s *Theory of Speech and Language*, described as “a presentation of the same problem from a somewhat different point of view.” Hjelmslev’s own *Omkring sprogteoriens grundlæggelse* appears as suggested further reading on “Types of Linguistic Structure,” while Jespersen’s *Modersmålets fonetik* and Jakobson’s *Kindersprache, Aphasie und allgemeine Lautgesetze* find their place under “Types of Linguistic Usage.” “Genetic Relationship” is represented by Meillet’s *Introduction à l’étude comparative des langues indo-européennes*, Schrijnen’s *Einführung in das Studium der indogermanischen Sprachwissenschaft*, and Charpentier’s *Jämförande indoeuropeisk språkvetenskap*. References under “Language Families” are to Lombard’s *Europas och den vita rasens språk*, to Kieckers’ *Die Sprachstämme der Erde*, and to *Les Langues du Monde*. Schrader’s *Sprachvergleichung und Urgeschichte* and Vilhelm Thomsen’s “Oldarisk kultur” (in the first volume of his *Samlede afhandlinger*) are the recommended readings on “Parent Languages.”

A few words remain to be added about the present version of the book. Besides the reading list just mentioned, I have eliminated another appendix that was more appropriate to the original edition than it would have been to this one – a short collection of selected terms and their definitions, together with some very brief biographical notes on linguists mentioned in the text. As translator, I have not thought it any part of my duty to conceal the fact that the book was originally addressed to a Danish audience: thus, the reader will find Latium still compared in size to Zealand (2700 square miles), and counterexamples introduced by an analogy with the legal situation on the Faeroes. At the same time, I have freely replaced or modified illustrative examples which, because of the knowledge of Danish that they presupposed or for other reasons, would have required the addition of distracting footnotes. In matters of terminology, I have deviated from the original in a few instances. I have used “relation” and “correlation” in the technical senses given to those words in the *Prolegomena*, simply because I found no other way of rendering, without risk of confusion, the less formal terms that Hjelmslev used in *Sproget*. I should also note particularly that the English text has expression elements “represented” by sounds or letters, where the Danish uses words that would normally be translated

“named” or “signified.” I can believe that Hjelmslev might, in English, have preferred “symbolized,” but, with some misgivings, I have settled on a relatively neutral substitute.

The final chapter, on languages of different degrees, is not included in the Danish edition. It did, however, originally constitute the conclusion of the book, as I discovered in 1967 when reviewing Hjelmslev’s papers at the gracious invitation of Mrs. Hjelmslev. Although passing years have obscured the history of this lost chapter, it would appear that Hjelmslev had at some time been persuaded that it was too difficult for a popular audience and that the book could be reduced to a more reasonable length by its omission. However that may be, it was preserved in fair copy, and it clearly represents the summation of all that has gone before. For these reasons I have not hesitated to include it here. The attentive reader may find it particularly helpful in dispelling the curious misconceptions that have arisen about Hjelmslev’s supposed “neglect of substance” and exclusive preoccupation with form.

Louis Hjelmslev authorized me to undertake the translation. I offer it to Vibeke Hjelmslev with my warmest gratitude.

F. J. W.

Berkeley

October 3, 1969

Prefatory Remarks

The science of language, or linguistics (from Latin *lingua* ‘a language’), has – like every other science – had its classical and its critical periods. In the classical periods there has been an established body, or structure, of doctrine, agreed on and respected by everyone, and scientific work has been concerned with details that could be fitted into the frames of that structure. In the critical periods, investigators have been occupied with the structure itself, which they have tried to establish on new and better foundations, reflecting a deeper and more penetrating appreciation of the nature of language.

Never, perhaps, has linguistics passed through a critical period of such profundity and scope as in our time. The very nature of language and its structural peculiarities are now being debated by linguists. The very basis of linguistics is being recast. This is not, of course, to be understood as meaning that linguists have ceased to be concerned with details. But even in their work on details they are constantly being faced with problems of foundations, to a much higher degree than ever before. This concern with foundations has already taken us so far from earlier notions that the new linguistics of our time can be contrasted to all former varieties, which can be grouped together under the common name of *classical linguistics*. Sooner or later the present crisis must necessarily lead to a new classicism. Perhaps we may even now say that we have come far enough to make out at least the contours of such a new classicism, but we have not yet attained it, and there is no agreement about the ways and means doing so.

Such a crisis is not to be deplored – quite the contrary. The critical periods are the refreshing and inspiring ones, proving that science is not stagnant but is constantly renewing itself. And the critical periods broaden our horizon and deepen our understanding of the things that have the most importance of all. Indeed, so great has been this effect of the present crisis that we may venture to assert that only now has linguistics succeeded in organizing itself as an independent science.

It is possible to be interested in language from many different points of view. The ancient Greeks and Romans and the mediaeval scholastics were interested in it mainly from the logical point of view; and it was in close association with Aristotelian logic that they founded a grammatical tradition still surviving in our school grammar and, with little change, even in most of our scientific grammar in its classical forms. In the eighteenth and nineteenth centuries, many scholars came to be interested in language principally from an historical and prehistorical point of view, with particular reference to cultural

history, so that the history and relationships of languages were seen as a reflexion of, and as a means of knowing, the history and relationships of peoples. In the nineteenth century arose the special science that called itself phonology or phonetics, interested in language from a physiological and physical point of view and attempting to describe the expression plane of language through a study of the movements and positions of the speech organs in the production of speech sounds and through a study of the sound waves that are so produced. There also arose in the nineteenth century a linguistic psychology, which saw in the psychological side of language, in language as a part of man's "mental life," the essential feature in the nature of language. And down to our own time, sociological considerations in ever-increasing measure have been brought to bear on language considered as a social institution.

Evidently, there is no lack of points of view from which language can be, and has been, studied. But none of these (and more could easily be added) provides the basis for an independent science of language; rather, language becomes an object of study, now for logic now for history, now for physiology, physics, psychology, or sociology. And it can be argued that despite all this many-sided study, one point of view about language has been neglected and, at that, the one that seems the most important and the most natural — the *linguistic* point of view. It should be possible to imagine a science that does not take language simply as a conglomerate of logical, historical, physiological, physical, psychological, and sociological factors, but first and foremost as an independent entity, an integral formation of a special kind. Only such an integral view can account for the fact that all these apparently so heterogeneous elements are able to come together in language. The linguistic point of view, taken as the central one, can determine the relative justification and the relative limitation of each of the particular points of view. Only when the logical, historical, physiological, physical, psychological, and sociological linguistics are complemented by a *linguistic* linguistics can an autonomous science of language be established. It is in a struggle to this end that our present crisis finds us.

Any exposition made today of the problem of language must necessarily reflect this situation — both for good and for ill. One advantage deriving from this critical debate is that we believe we have reached a clear understanding of many things that formerly either went simply unnoticed or were viewed in a quite different light. One disadvantage is the multiplicity of discordant voices echoing round the author as he writes. It would hardly be possible, in a popular survey like this, to allow all these many voices to be heard and to introduce them directly into the critical debate. And it would serve little purpose,

as well. There are many – indeed, many important – points at which one dares to break into the debate, strike out, and say: this is the way things are, and that is not the way things are. And it would be time-wasting and confusing to explain that there are people who have thought, or who perhaps still think, that things are different. Many positions have been held without any justification in the earlier, subjective period of linguistics. Through the present crisis, justification, argumentation, and proof are demanded, and opinions offered without justification need be neither reported nor refuted: they will die of their own accord.

The reader must realize, however, that it is hardly possible to set up a complete system at the present time. In many respects, the exposition that follows must aim at posing problems rather than solving them. This book, then, is designed as an introduction to the *problem* of language as, *in the author's judgment*, it stands today.

The reader should also understand two things:

1. Our presentation is *not subjective*. To be sure, a strong personal element is unavoidable in respect of the ends and means of research as well as in the way the problems are conceived, but the views that are offered here are not subjective: they are objective, being based on arguments. They represent, not mere belief, but substantiated reasoning.

2. Our presentation is *not exclusive*. It attempts to do justice to all sides of linguistics. The fact that a science is adopting new points of view does not mean, as laymen may perhaps occasionally believe, that for that reason it rejects all previously obtained results as dispensable or even regrettable. Although Tycho Brahe erroneously assumed that the sun revolves around the earth, it does not follow that his observations, discoveries and results have to be abandoned, but only that they have to be *reinterpreted*, converted according to the requirements of the newer point of view, and understood in a different way. The same holds true of those fields within classical linguistics where true observations and discoveries were made. In science it is quite possible to speak of permanent result – although hardly of permanent points of view. Nineteenth-century classical linguistics did achieve permanent results as regards the genetic relationship of languages, results that constitute an essential side of linguistics. They will be presented here, however, in adjustment to new points of view, and they will be looked at somewhat differently from the way in which they perhaps were seen at the time of their discovery, so that the following account will differ in its principles from what is usually found in the handbooks on the subject.

Such an adjusted presentation of the main results of nineteenth-century linguistics will be our starting point, because the field to which they belong – the study of genetic relationship of languages – is still the best cultivated within the whole science of linguistics and because it is the only one for which classical linguistics produced what we can call – at least after reinterpretation – an exact method. This cornerstone of nineteenth century classicism is also what gave rise to the critical approach of modern linguistics, and it will be used accordingly to introduce the reader to the modern framework within which problems of linguistic structure and linguistic change are treated.

Linguistic Function

We gain insight into language, acquaintance with it, understanding or comprehension of it in the same way as we gain insight into other objects – through a description. And to describe an object can mean only one thing: to give an account of the relationships into which it enters or which enter into it. Such relationships, or dependences, registered by a scientific description, we shall call *functions*. Thus we can describe a given object in two ways: (1) by dividing it into parts with mutual function, or *analyzing it*; (2) by placing it within a whole whose parts have mutual function, or *synthesizing it*. In the first case, the object is itself viewed as a functional whole; in the second, it is viewed as a part of a larger functional whole.

Accordingly, when the object we wish to describe is a language, we may

1. on the one hand, analyze the language into parts with mutual function; this is done in that kind of linguistics called *grammar*, which views the individual language (more precisely, the individual *état de langue*) as a closed totality and describes it by analyzing it and accounting for the functions between its parts;

2. on the other hand, place the language within a larger whole, i.e. view it as part of a *class of languages* and then describe the class by analyzing it and accounting for the functions between its parts.

The connexions between languages belonging to one and the same class of languages – or, in other words, the functions that establish a linguistic class – constitute *linguistic relationship*, of which we distinguish two kinds:

1. *genetic* relationship, which is found between languages belonging to the same *language family* and which points to a common origin for them;

2. *typological* relationship, which is found between languages belonging to the same *language type*; typological relationship does not point to a common origin, but is based on an agreement in structural features that is conditioned by the general possibilities of language.

In what follows we shall come to see what sorts of functions establish these two kinds of language classes, the language families and the language types. We shall also have occasion to determine some of the most important internal functions of the *état de langue*. Actually, of course, the analysis of the individual *état de langue* is a prerequisite for the classification of languages, but the chief problem of linguistics will always be, not the individual *état de langue*, but the relationship between different stages of a single

language and between different languages, their similarities and their differences. This chief problem will therefore be attacked directly, while questions concerning the état de langue will be discussed as they arise and require elucidation in the course of the exposition.

Genetic Relationship of Languages

Our own language belongs to the family usually called *Indo-European* (so named because it includes languages of both India and Europe). In what follows, a few examples of words from this family will be given to illustrate genetic relationship. If these examples have been chosen from the Indo-European family and not, as they might just as well have been, from some other, this is not only because the Indo-European family includes particularly well known languages (though whether the languages cited are known to the reader or not will be of no importance for his understanding the argument) and because our language belongs to it, but because, of all language families, it has been the most thoroughly studied and has been traced through the longest period of evolution and the broadest geographical diffusion. We know Indo-European languages over a period of 3500 years, and Indo-European languages are now spoken by 1.4 billion people, about half the world's population. By choosing this family as an example, we can demonstrate that our method is subject to no limitations of time or space.

On the other hand, we shall wish to choose our examples from among the many available so as to present our material in the simplest possible form and so as to avoid unnecessary complications that might at first obscure the general picture. In order to satisfy those two requirements – demonstration over a great extent of time and space, and simplicity of exposition – we shall choose our examples from certain Indo-European languages which (for reasons that can be discussed later) are particularly well suited to the purpose. The choice of languages will vary slightly from one example to another, but in the main we shall be using the languages listed in the following paragraph.

Among the languages having closest genetic relationship to our own⁴⁶, we shall usually choose *Gothic* (Goth.), which was spoken from about the year 200 to about the year 500 in large areas of Europe, from Spain in the West to the Crimea in the East, and into which the Bible was translated in the fourth century, in what is now Bulgarian territory. Only when a word is not attested in Gothic (i.e., not found in the extant Gothic texts) will we choose other closely related languages instead: *Old Norse* (ON), the language of the Icelandic sagas and the Norwegian Eddas; *Old High German* (OHG); or *Old English* (OE). To represent the Celtic languages we shall use either *Irish* (Ir.) or

⁴⁶ N. T. Hjelmslev refere-se aqui ao dinamarquês, cuja proximidade com o gótico justifica a escolha. Na tradução inglesa, não há esse problema, visto que o inglês possui, também, uma proximidade com o gótico.

Welsh. Further, we shall cite *Latin* (Lat.), *Greek* (Gr.), *Lithuanian* (Lith.), *Old Church Slavonic* (OCS), *Armenian* (Arm.), and *Old Indic* (OI) – and in some cases *Hittite* (Hitt.), an ancient language of Asia Minor that is known from a large number of cuneiform inscriptions, and *Tokharian* (Tokh.), a dead language preserved in manuscripts found in Chinese Turkestan. Further information about these and other related languages is not needed at the moment, but will be given in the chapter on language families.

We shall now first look at the words *brother*, *mother*, and *father*. Our word *brother* is here cited from all the languages named above, excepting only Hittite:

Goth. <i>broþar</i>	Gr. <i>phrátōr, phrátēr</i>	Arm. <i>etbayr</i>
Ir. <i>bráthir</i>	Lith. <i>broter-ė̃lis</i>	OI. <i>bhrátā</i>
Lat. <i>frāter</i> (cf. Eng. fraternize)	OCS <i>bratrъ</i>	Tokh. <i>pracarə</i>

Note:

1. So far as the individual letters of the transcriptions are concerned, it is enough for our purpose if the reader is able to recognize them when he meets them again later. Beyond that, he need not to be concerned with these oddities or try to imagine the pronunciation of the words cited. The following notes may serve to satisfy a natural curiosity, however: *þ* is pronounced like *th* in Eng. *thing*; a horizontal stroke over a vowel indicates that it is long; *ł* is pronounced like *l* with the back of the tongue raised (“dark” *l*); *ə* is pronounced like *a* in Eng. *sofa*; OCS *ъ* is an extra-short vowel; Lith. *ė* is a long close *e* (like the first vowel in Ger. *Seele*), and the circumflex (*~*) indicates a special intonation; Tokh. *c* is pronounced like *ch* in Eng. *child*.

2. The Lithuanian form is a hypocoristic and therefore supplied with a suffix *-ė̃lis*, which we do not find in the words cited from the other languages. Another word for ‘brother’ in Lithuanian is *brólis* (where the acute accent indicates a special intonation different from the circumflex), which, although related to our word *brother*, is not identical with it.

3. The word means ‘brother’ in all the languages except literary Greek, where it has the special meaning ‘(fellow-)clansman’, but we know that the same word in the Ionic dialect (where it has the slightly different form *phrétēr*) did mean ‘brother’. In common Greek, ‘brother’ is *adelphós*, which is a different word. A ‘clan’, i.e. ‘brotherhood’, in

Greek is *phrātría*, a derivative of the word *brother* which we also find in OCS *bratrija* and OI *bhrātryam*.

Our word *mother* is not preserved in Gothic (in the Gothic Bible, ‘mother’ appears as *aiþei*, which must be a different word), but we can cite:

ON <i>móðir</i>	Gr. <i>mātēr</i> (Ático <i>métēr</i>)	Arm. <i>mayr</i>
Ir. <i>máthir</i>	Lith. <i>mótė, motė̃</i> ‘woman’, <i>mótina</i> ‘mother’	OI <i>mātá</i>
Lat. <i>māter</i>	OCS <i>mati</i>	Tokh. <i>mācarə</i>

(Unless otherwise noted, the word in each language means ‘mother’.
Pronunciation: *ð* like *th* in Eng. *there*; Ir. *th* like *th* in Eng. *thing*.)

Our word *father* is not found in Lithuanian or Old Church Slavic the Lithuanian word for ‘father’ is *tėvas*, which may perhaps be very distantly related to our word *father* but in any case is not identical with it; the Old Church Slavic is *отць*, a completely different word. Nevertheless, we can cite the following words, all meaning ‘father’:

Goth. <i>fadar</i>	Tokh. <i>pācarə</i>
Ir. <i>athir</i>	Gr. <i>patér</i>
Lat. <i>pater</i> (cf. Eng. <i>paternity</i>)	Arm. <i>hayr</i>
	OI <i>pitá</i>

We shall now turn our attention to the *expression elements* composing these words in the different languages. These expression elements are represented in writing by letters, and in spoken language by speech sounds. (In some cases we use a combination of letters to represent one expression element: Ir. *th*, Gr. *ph*, and OI *bh* will be counted as representing one expression element each.) If we disregard the accents (which, of course, are also expression elements), the Greek words *phrātēr*, *mātēr*, *e patér*, for example, will be found to contain six, five, and five expression elements respectively.

Now it can be shown that between each expression element (EE) in one language and each EE in another, genetically related language there is a fixed dependence or relationship, or, as we have agreed to call it, a function. It is, in fact, on this function

between the EE's of the different languages, on this *element-function*, that genetic relationship between languages rests. Incidentally, it is also due to this element-function that we often find, in such examples as the foregoing, a resemblance between words in different languages (although this need not be present – it would hardly be said that the Armenian words bear a very great “resemblance” to the others).

An especially simple case is the first EE in the word *mother*, which appears as *m* in all the languages we have cited. Now what is interesting and crucial is the possibility of establishing a rule that each time we have an *m* at the beginning of a syllable in a word in one of these languages, we have an *m* in the same word in all the other languages, provided the word is found in them. To be sure, such a rule is not without restrictions, but these restrictions can be stated quite definitely and kept separate. We shall return to this question later, in our discussion of sign formation. First of all, we shall demonstrate the correctness of our rule by means of some examples:

mouse: ON *mús*, Lat. *mūs*, Gr. *mûs*, Arm. *mukn*, OI *mūh*

moon, with its derivative, *month*: Goth. *mena* ‘moon’, *menops* ‘month’; Ir. *mí* ‘month’; Lat. *mēnsis* ‘month’; Gr. *mén* ‘month’, *ménē* ‘moon’; Lith. *mėnuo* ‘moon’, *mėnesis* ‘month’; OCS *měsęcь* ‘moon, month’; Arm. *amis* ‘month’; OI *māh* ‘moon, month’. (OCS *ě* pronounced approximately like *a* in Eng. *man*; *ę* like Fr. *fin*; *c* like *ts*; *ь* an extra-short vowel different from *ъ*; OI *h* is an ordinary *h* originating from *s*.)

me: Goth. *mi-k*, Welsh *mi*, Lat. *mē*, Gr. *me*, *emé*, *emé-ge*, Lith. *manė*, OCS *mę*, IDA *mā*.

murder: Goth. *maurþr* ‘murder’; Lat. *mors* ‘death’; Gr. *mortós* ‘mortal’; Lith. *mirtis* ‘death’; OCS *сѣ-мрътъ* ‘death’; Arm. *mard* ‘a man’ (a mortal); OI *mṛtiḥ* ‘death’. (*r* indicates a syllabic *r*).

‘winter’: ON *gimbr* ‘yearling (one-winter-old) ewe lamb’ (cf. dialectal Eng. *gimmer*). Ir. *gemred* ‘winter’; Lat. *hiems* ‘winter’; Gr. *kheima* ‘winter frost’; Lit. *žiemà* ‘winter’; OCS *zima* ‘winter’; Arm. *jmeṛn* ‘inverno’; OI *himáh* ‘frost, snow’, *himā* ‘winter’, *héman* ‘in winter’; Hitt. *kimanza* ‘winter’. (*ž* is pronounced like *j* in Fr. *jour*; Hitt. *z=ts*; Arm. *j=dz*; Arm. *ř* is a strongly trilled apical *r*.)

tame: Goth. *ga-tamjan* ‘to tame’; ON *tamr* ‘tame’, *temja* ‘to tame’; Ir. *damnaim* ‘I tie, I tame’; Lat. *domāre* ‘to tame’; Gr. *damân* ‘to tame’; OI *dāmyati* ‘he tames, is tame’.

When *m* stands at the end of a syllable, it again appears as *m* in most of the languages, but in Old Church Slavonic it combines with the preceding vowel to produce a

nasal vowel. Starting from the meaning ‘serrate’, we can identify our word *comb* ON *kambr* ‘comb, crest, jagged beach’, with Gr. *gómphos* ‘bolt dowel’, OI *jámbhaḥ* (*j* pronounced as in English), Lith. *žam̃bas* ‘uneven edge’. In Old Church Slavic the same word is *zqbъ* ‘tooth’. Such will be the case in all examples where *m* stands at the end of an internal syllable. The relationship is again different when the *m* stands at the end of a word: then, in Lithuanian we find a nasal vowel (no longer pronounced as a nasal in the standard language); in Old Church Slavic, we sometimes find nasalization of the vowel and sometimes zero, depending on certain prosodic conditions; in Greek we find *-n*; and in Gothic and Old Norse we find zero. Thus the word that we have just cited in the nominative case will, in the accusative (Indo-European ending *-m*), assume the following forms: OI *jámbham*, Lith. *zam̃bq*, OCS *zqbъ*, Gr. *gómphon*, ON *kamb*. If the word had existed in Latin, the accusative form would have ended in *-m*, like other accusatives.

We thus find a constant function between the EE’s of these languages, such that the same correspondences always appear under the same conditions. Using the symbol φ to designate this constant dependence or function, we can state that under the first condition – syllable initial position – we have:

$$\text{Goth. } m \varphi \text{ Celt. } m \varphi \text{ Lat. } m \varphi \text{ Lith. } m \varphi \text{ OCS } m \varphi \text{ Arm. } m \varphi \text{ OI } m$$

But it would be rather cumbersome to write or to say all this every time we wished to refer to such an element-function between all the Indo-European languages, and for that reason, we use a more convenient system in which a single function through all the languages is designated by a single sign. Naturally, this sign – like all scientific symbols – is chosen quite arbitrarily. We could, if we wished, agree to use *x* or *y* or *z* to refer to the whole long functional formula above. But the signs used in scientific formulæ are chosen not only arbitrarily, but also appropriately, and so, from purely practical considerations, we shall choose a formula that will directly recall to us something of what it is intended to designate. In the example before us, there can be no doubt that the simplest thing will be to choose the letter *m*, before which we shall set an asterisk to indicate that it is a formula for an element-function. Thus, instead of saying each time “agreement between Gothic *m*, Celtic *m*, Latin *m*, etc., at the beginning of a syllable,” we shall say “Indo-European *m*” and write “IE **m*.”

Besides arbitrariness and appropriateness, scientific formulæ have the additional property of simplicity. When we set about studying all the element-functions between the

Indo-European languages and face the task of registering them, we shall prefer to use as few symbols as possible. To continue with our same example, we find that the conditions under which the function is different (syllable-final and word-final position) and the conditions under which our first function holds (syllable-initial position) are mutually exclusive. This being so, we can allow ourselves to use the same symbol in all these cases without running the risk of confusion. And since there are still languages where the EE concerned is represented by *m* in syllable-final and word-final positions, we can still use the symbol **m* and set up the equations

$$\text{IE } *m = \text{Goth. ON } m \varphi \text{ Gr. } m \varphi \text{ Lith. } m \varphi \text{ OCS, OI } m$$

in syllable-final position within the word, and

$$\text{IE } *-m = \text{Goth. ON } \emptyset \varphi \text{ Lat. } -m \varphi \text{ Gr. } -n \varphi \text{ Lith., } \varphi \text{ OCS } \emptyset (.) \varphi \text{ OI } -m.$$

In the latter formula we have even been able to state the condition (word-final position) simply by placing a hyphen before the *m*.

We do the same for all EE's, one by one, remembering to make do with the fewest possible Indo-European formulæ and therefore to bring together under a single formula as many different conditions as possible. The method is purely mechanical and easy to apply so long as the lexical material is available. Thus, the final element in *brother*, *mother*, and *father* can be identified with the second element in *brother* and denoted by the formula **r*. The reader can easily see from the given examples – to which many others could be added – that most of the languages under most conditions have a corresponding *r*, although in word-final position Lithuanian and Old Indic have zero. This zero will, of course, be in turn replaced by *r* when we change the conditions – that is, when we introduce into the same word, so far as that is possible, the conditions under which the language concerned has a corresponding *r*: in the accusative, Old Indic has *brhátar-am*, *mātár-am*, *pítár-am* and Lithuanian has *móter-į*; cf. also Lith. *broter-ėlis*.

The same examples (and, again, many more could be added) make it immediately obvious that the third EE in *brother* and the second EE in *mother* are identical and can be formulated as

IE * \bar{a} = Goth. *o*, ON *ó* φ Ir. *á* φ Lat. *ā* φ Gr. *ā* (\bar{e}) φ Lith. *o* φ OCS *a* φ Arm. *a* φ OI *ā* φ Tokh. *ā*
(*a*).

IE * \bar{a} can also appear at the end of a word, but under this condition it corresponds to other functional formulæ. We have already seen examples of this: since * \bar{a} is a feminine nominative ending (cf. the Latin first declension, in *-a*), it is found in a large number of words, including those we have cited above as OI *himā* ‘winter’, Lith. *žiemà*, OCS *zima*, and Gr. *ménē* ‘moon’, which would be *ménā* in dialects other than Attic and Ionic. On this basis we can conclude:

IE * \bar{a} = Lat. *-a* φ Gr. *-ā* ($-\bar{e}$) Lith. *-a* φ OCS *-a* φ OI *-ā*

The second EE in *father* reveals the function

Goth. *a* φ Ir. *a* φ Lat. *a* φ Gr. *a* φ Arm. *a* φ OI *i* φ Tokh. *ā*,

Which we sum up in the symbol IE * Λ (reserving **a* for the cases where OI *a* corresponds to *a* of the other languages, as in OI *ájati* ‘drives’, arm. *acem* ‘I lead, bring’, Gr. *ágein* ‘to lead’, Lat. *agere* ‘to drive, lead’, Ir. *agat* ‘let them lead’, ON *aka* ‘to drive’). This correspondence holds only for the first syllable of a word; other syllables show other element-functions corresponding to IE * Λ .

Let us conclude this sampling of element-functions by taking a closer look at the first EE in the word *father* and the third EE from the end in the words *brother*, *mother* and *father*. For the former we shall write IE * p , and for the latter, IE * t .

The word *father* displays the function

IE * p = Goth. *f* φ Ir. \emptyset φ Lat. *p* φ Gr. *p* φ Arm. *h* φ OI *p* φ Tokh. *p*,

A function that we find in an enormous number of other examples, some of which also show us that the Hittite correspondence is *p*. We shall cite a few of these:

feather: ON *fjoðr*, OHG *fedara* ‘feather’; OHG *fed(a)rah* ‘wing’; Ir. *én* ‘bird’; Old Welsh *eterin* ‘bird’; Lat. *petere* ‘to seek, strive for, make for, travel to’; Gr. *péteshai* ‘fly, rush’; OI *pátati* ‘it flies’, *pátram* ‘wing’; Hitt. *pítaizi* ‘fless’, *pítar* ‘wing’.

fish: Goth. *fisks*, Middle Ir. *iasc*, Lat. *piscis* (in Slavic, Pol. *piskorz* ‘loach’, Russ. *piskar* ‘gudgeon’). (Lat. *c* is pronounced like *k*.)

ME *fee* ‘cattle, property’: Got. *faihu* ‘property, money’; ON *fé* ‘cattle, money, property’; Lat. *pecu*, *pecus* ‘cattle’, *pecúnia* ‘wealth, money’ (cf. Eng. *pecuniary*); OI *pásu*, *paśú*, *paśúh* ‘cattle’.

Dan. (*i*)*ffor* ‘last year’: ON *fforðr*, Gr. (Doric) *péruti*, Lith. *pérmai*, Arm. *heru*, OI *parút*.

The conditions for this function turn out to be quite restricted, when we consider all the languages participating in it. In Armenian, for example, we do not find *h*, but zero, when **p* stands word-initially before IE **o* (thus, Arm. *otn* corresponds to Eng. *foot*, Goth. *fotus*, Gr. *poús*). But under most conditions this function is found at the beginning of words. In the group **sp*, we find *p* in all the languages, including Gothic, so, for example ON, *spá* ‘to foretell’, OHG *spehōn* ‘to spy’, as compared with Lat. *specere*, *spectāre*, *inspicere* ‘to inspect’, *haru-spex* ‘augur (who inspects the victim’s entrails for omens)’. But within the word, **p* corresponds to quite different functions, as we shall see shortly.

The EE function Goth. ON *f* \varnothing Lat. Gr. *p* was first recognized by Rasmus Rask, in 1814 (in his prize essay entitled *Investigation on the Origin of the Old Norse or Icelandic Language*), in connexion with a whole set of other EE functions holding for word-initial position in the same languages, like Goth. ON *t* \varnothing Lat. Gr. *d*, of which we have seen an example in the word *tame*. For this latter function (and corresponding functions under other conditions) we establish the formula IE **d*, and from our example of *tame*, supplemented by appropriate material from Lithuanian, Old Church Slavic, and Armenian, we can read off the following function:

IE **d* = Goth. ON *t* \varnothing Ir. *d* \varnothing Lat. *d* \varnothing Gr. *d* \varnothing Lith. *d* \varnothing OCS *d* \varnothing Arm. *t* \varnothing OI *d*.

We find good examples for this function in the numerals *two* and *ten*:

Goth. *twai*, Ir. *dáu*, Lat. *duo*, Gr. *dúo*, Lith. *dù*, OCS *dъva*, OI *dvau*.

Goth. *taihun*, Ir. *deich*, Lat. *decem*, Gr. *déka*, Lith. *dėšimt*, OCS *desęť*, Arm. *tasn*, OI *dása*. (*ś* is pronounced like Eng. *sh*.)

The set of word-initial EE functions recognized by Rask also includes the function for **t*: Goth. ON *þ* \varnothing Lat. Gr. *t*, or, more fully:

Goth. ON *þ* φ Celt. *t* φ Lat. *t* φ Gr. *t* φ Lith. *t* φ OCS *t* φ Arm. *th* φ OI *t*,

as exemplified by:

thirst: Goth. *þaursus* ‘dry’, *þaurstei* ‘thirst’, *þaursjan* ‘to thirst’, *ga-þairsan*, *ga-þaursnan* ‘to dry up’; Ir. *tart* ‘thirst’; Lat. *torridus* ‘seco’, *torrēre* ‘to dry up, roast, bake’; Gr. *térsesthai* ‘to become dry’, *tersáinein* ‘to dry’, *trasiá*, *tarsiá*, *tarsos* ‘drying crate’; Arm. *thar* ‘rod for drying grapes’; OI *tr̥syati* ‘is thirsty’, *tarśáh* ‘thirst’, *tr̥štáh* ‘dry, hoarse’. (*š* and *t̥* pronounced with the tip of the tongue turned back to the roof of the mouth.)

thin: ON *þunnr*, Ir. *tanae*, Lat. *tenuis*, Gr. *tanaós* ‘long’, OCS *тънькъ* ‘thin, slender’, OI *tanúh*, *tánukaḥ*. (The meaning is ‘thin’ unless otherwise indicated.)

This functional formula holds for word-initial position before a vowel. Under other conditions, word-initial **t* will have other representations in some of the languages, but in Gothic and Old Norse it is always represented by *þ*. In the group **st*, Gothic and Old Norse have *t*, just as they have *p* for **p* in **sp*, compare Goth. *steigan* ‘to climb up’ and Gr. *steíkhein* ‘to walk, march’.

But as we can see by the third EE from the end in *brother*, *mother*, and *father*, there are functions for **t* within the word that differ from what we find in word-initial position; moreover, we observe that Gothic has different representatives in *broþar* and *fadar*, corresponding to a single representative in the other languages. The conditions must therefore be different in the two words. If we collect the examples for this contrast, we find that what decides whether we shall find *p* or *d* in Gothic is the place of the accent in the corresponding Old Indic word. If the word accent in Old Indic immediately precedes the *t*, Gothic has *þ*, otherwise *d*; thus *broþar* corresponds to OI *brátā*, but *fadar* to OI *pitá*. This relationship between Gothic consonants and Indic accent was discovered by the Danish linguist Karl Verner in 1876. We have already stated (p. 27) that the accents must be counted among the expression elements; Verner’s law implies, of course, that the Old Indic accents must be included in our Indo-European formulæ.

It turns out, as a matter of fact, that accent relationships (or, at least, phenomena that appear to us most directly in the form of accent relationships) play a role in the EE functions in many other respects. It has been shown that the Greek opposition between circumflex (ˆ) and acute (ˊ) accentuation of the accented syllable is also found or reflected

in many other Indo-European languages (most clearly in Lithuanian, where $\tilde{}$, on the one hand, and $\acute{}$ or $\grave{}$, on the other, are distributed in final syllables as are $\hat{}$ and $\acute{}$ in Greek). The formula given above (p. 30) for IE $*-\bar{a}$ holds only when the $*-\bar{a}$ stands, in Greek and Lithuanian, in the kind of syllable that takes the acute when accented (and this is precisely the case for the ending of the feminine nominative). When $*-\bar{a}$ stands in the kind of syllable that takes the circumflex when accented in Greek and Lithuanian, we obtain another function. In such cases, Lithuanian has *o*, not *a*, corresponding to IE $*-\bar{a}$, as in the feminine genitive ending represented by Gr. $-\bar{a}s$ ($-\bar{e}s$) and Lith. *-os*.

We have thus found a relationship between the Indo-European languages under consideration such that a given EE in one language, in the environment of other given EE's and in a given place in the word, has a constant correspondence to a given EE in each of the other languages. And this relationship can be demonstrated for all EE's in all Indo-European languages. Thus the whole EE-system in each of the languages stands in a constant relationship to the whole EE-system in each of the other languages. It is this constant correspondence that we call the EE-function or *element-function* (provided we agree to understand this latter term as referring to *expression* elements).

The principle behind this observation can be generalized. It is universal: wherever there is genetic relationship between languages, there is element-function between their systems. This is true of all language families, far and near, small and large. In those instances where it has been possible to establish genetic relationship between American Indian languages, the very same method has been used. When we establish, or try to establish, genetic relationship between Indo-European and other languages, it is again a matter of establishing element-function. And element-function is also used to establish the kind of closer relationships found within subfamilies. For a large language family like Indo-European contains subfamilies for which we can establish specific element-functions and set up corresponding formulæ. In other words, there are language families within language families, or language families of different degrees.

One subfamily within Indo-European is *Italic*, which includes Latin and the modern *Romance* languages (Italian, Spanish, French, Rumanian, etc.). Here is one example of an element-function between the Romance languages:

$*ct = \text{It. } tt \varphi \text{ Es. } ch \text{ (pronounced like Eng. } ch) \varphi \text{ Fr. } it \varphi \text{ Rom. } pt,$

as in:

It. <i>fatto</i>	Sp. <i>hecho</i>	Fr. <i>fait</i>	Rum. <i>fapt</i>	‘done’
It. <i>latte</i>	Sp. <i>leche</i>	Fr. <i>lait</i>	Rum. <i>lapte</i>	‘milk’
It. <i>notte</i>	Sp. <i>noche</i>	Fr. <i>nuit</i>	Rum. <i>noapte</i>	‘night’

Just as we called our formulæ Indo-European when we were comparing the Indo-European languages, so the Romanists give their formulæ a name, calling them Common Romance Proto-Romance, or Vulgar Latin. From considerations of appropriateness, these are chosen so as to resemble as much as possible the corresponding EE’s in Latin as it has come down to us. But they are just as arbitrary as the Indo-European formulæ, and there are many instances where it would be impossible to bring them into agreement with the EE’s in the Latin words.

Another subfamily within Indo-European is *Germanic* (Gmc.), which includes Gothic, Old Norse and the other Scandinavian languages, and the several stages of German and English. Its formulæ are called Common Germanic or Proto-Germanic, and it turns out to be appropriate to choose them so that in many (but not all) cases they will resemble the EE’s of Gothic. Thus, for example, the Germanist has no need of starting from IE **t*, but can set up a Gmc. **þ* and show which functions between the Germanic languages correspond to it under different conditions. For the condition illustrated by the examples given above we obtain the following function:

Gmc. **þ* = Goth. *þ* \varnothing ON *þ* \varnothing Modern Scand. (Dan., Swed., Nor.) *t* \varnothing OE *þ* \varnothing Modern Eng. *th*
(voiceless, as in *thin*) \varnothing HG *d*.

Citing the languages in this order, and returning to our old examples from this new point of view, we find:

Goth. *þaurusus*, etc.; ON *þurr* ‘dry’, *þerra*, *þorna* ‘to dry’, *þorsti* ‘thirst’, *þyrstr* ‘thirsty’, Dan. *tør* ‘dry’, *tørre* ‘to dry’, *tørst* ‘thirst’; OE *þyrre* ‘dry’, *þurst* ‘thirst’; Modern Eng. *thirst*; OHG *durri* ‘dry’ *derren*, *dorrên* ‘to dry’, *durst* ‘thirst’; Modern HG *dürr* ‘dry’, *dorren* ‘to be dried’, *dörren* ‘to dry’, *durst* ‘thirst’.

ON *þurnnr* ‘thin’, Dan. *tynd*, Modern Eng. *thin*, OHG *dunni*, Modern HG *dünn*.

In corresponding fashion, we can set up Gmc. **t*, from the Germanic point of view, for the examples used above to set up IE **d*:

Gmc. **t* = Goth. *t* \varnothing ON *t* \varnothing Modern Scand. *t* \varnothing OE *t* \varnothing Modern Eng. *t* \varnothing HG *z*.

as in

Goth. *ga-tamjan*, ON *tamr*, *temja*, Dan. *tam*, *tamme*, OE *tam*, *temian*, Modern Eng. *tame*, OHG *zam*, *zamôn*, *zemma*, Modern HG *zahn*, *zähmen*.

Goth. *twai*, ON *tveir*, Dan. *to*, Modern Eng. *two*, OHG *zwêne*, Modern HG *zwei*.

Goth. *taihun*, ON *tíu*, Dan. *ti*, OE *tíen*, Modern Eng. *ten*, OHG *zehan*, Modern HG *zehn*.

The EE-functions HG *d* \varnothing Goth. ON *þ* and HG *z* \varnothing Goth. ON *t* were recognized in 1814 by Rask, in the essay already mentioned, in connexion with some other EE-functions for the same languages. Just as Gmc. *f* = IE **p* and Gmc. *þ* = IE **t* are particularly striking and characteristic functions for Germanic in its relationship to other Indo-European languages, so HG *d* = Gmc. **þ* and HG *z* = Gmc. **t* are particularly striking and characteristic functions for High German in its relationship to the other Germanic languages. (Low German has *th* (later *d*) and *t* respectively: in Old Saxon – the old form of Low German – ‘thin’ is *thunni* and ‘ten’ is *tehan*. This is one of the most immediately striking differences between the two languages of Germany, High German and Low German.) These characteristic High German functions also include *t* = Gmc. **d* (Goth. *d*), so that Verner’s opposition of Goth. *broþar* and *fadar* takes the form in High German of an opposition between *d* and *t*: *Bruder*, *Vater*.

Of course, a single EE in one language can correspond to two or more different EE’s in the related languages. We have seen above (p. 31) both OI *i* and OI *a* in the first syllable of words as corresponding EE’s to *a* in the other languages, where *a* therefore equals IE **A* in some words and IE **a* in others. As we can see, a distinction found in one language may be obliterated in another. Thus, Verner’s distinction between Goth. *broþar* and *fadar*, Modern HG *Bruder* and *Vater*, which is also found in OE *bróþor*, *fæder*, is obliterated both in Scandinavian (Icelandic *bróðir*, *faðir*, Dan. *bróder*, *fader*) and in Modern English (*brother*, *father*) since these languages have in this case only one EE corresponding to both Gmc. **þ* and Gmc. **d*. Another such case to be found in our examples is that of Dan. (Modern Scand.) *t*, which in some words (like *tør* ‘dry’ and *tynd*

‘thin’) corresponds to Gmc. **þ*, IE **t*, and in others (like *tam* ‘tame’, *to* ‘two’, and *ti* ‘ten’) to Gmc. **t*, IE **d*. Thus, genetically related languages by no means need have the same of EE’s or the same system of EE’s.

The subfamilies of which we have seen examples stand within the great Indo-European language family as states within a state. They are established according to the very same principle as the great family, but independently of it. The Germanist, or Romanist, can work within a closed system in which the Germanic, or Romance, formulæ are adequate tools, without having to look beyond the limits of the subfamily to the other members of the larger family. This fact determines a useful division of labor among the linguists engaged in the study of genetic relationship. Each language family, however small, as well as each part of a language family, is like a microcosm, organized in exactly the same way as the larger families. For example, we could consider the Scandinavian languages apart from the rest and establish a set of common formulæ just for them. We could restrict our domain even further, taking the East Scandinavian languages (Danish and Swedish) by themselves and treating them in quite the same way. And we could even take the Danish dialects by themselves and show how they have mutual element-functions that can be accounted for by a Common Danish system of formulæ. From this point of view, each domain is sufficient to itself. By selecting certain element-functions as peculiarly characteristic of a given language group, a given language, or a given dialect area (like the consonantal functions that we found to be specifically peculiar to High German, opposing it to all the other Germanic languages), we are led to distinguish language families of different degrees or of different size. And generally it is, of course, most expedient to stay within as narrowly restricted a family as possible when treating its internal problems. The Danish dialects are accounted for from Common Danish, Danish from East Scandinavian, and East Scandinavian from Scandinavian (or Norse). Further, if we follow the classification of the Germanic languages proposed on pp. 70-71⁴⁷, Scandinavian will be accounted for from East Germanic, East Germanic from Germanic, Germanic from Indo-European, and Indo-European from common formulæ for functions between Indo-European and non-Indo-European languages. It is this continuous gradation that determined the choice of languages (p. 25) to be used in exhibiting some of the Indo-European element-functions. We could, of course, compare Danish directly with Latin, or French with Old Indic, but such comparisons introduce complications that can be

⁴⁷ N.T.: Paginação da versão em inglês. Paginação nesta dissertação ainda a definir.

avoided if we first choose a Germanic language that is closer to Common Germanic than Danish, or an Italic language that is closer to Common Italic than French. Without an intermediate link, we should face the complications of having to account for all the instances where Danish obliterates a distinction found in Common Germanic or in Gothic, or where Danish introduces a difference that is not found there, or, correspondingly, where French obliterates a distinction found in Common Italic or in Latin, or introduces a distinction that is not found there.

Thus, for every language family, large or small, the linguist uses as his tools a set or system of common formulæ in which he summarizes the observed element-functions in the simplest and most economical way. Each language family can be accounted for from its system, and each period in the history of linguistic research has had its own system for each family, since further simplifications are often possible. For example, in the course of the last hundred-odd years at least three clearly different systems, each with its own nuances, have been set up for Indo-European. The third one is even now in the course of formation, and we shall discuss it a little later. In this section we have provisionally kept to a methodological exposition of the system arrived at by classical linguistics, that is to say, the second of the three. In summary, this system has the following aspects:

1. *Consonants*, i.e. entities that do not by themselves establish a syllable. Of these, we set up a minimum of $5 \times 4 + 3 = 23$:

<i>*p</i>	<i>*ph</i>	<i>*b</i>	<i>*bh</i>
<i>*t</i>	<i>*th</i>	<i>*d</i>	<i>*dh</i>
<i>*k₁</i>	<i>*k₁h</i>	<i>*g₁</i>	<i>*g₁h</i>
<i>*k₂</i>	<i>*k₂h</i>	<i>*g₂</i>	<i>*g₂h</i>
<i>*k₃</i>	<i>*k₃h</i>	<i>*g₃</i>	<i>*g₃h</i>
<i>*s</i>	<i>*p</i>	<i>*j</i>	

2. *Vowels*, i.e. entities that by themselves establish a syllable. Of these, we set up a minimum of $2 \times 3 = 6$:

<i>*a</i>	<i>*e</i>	<i>*o</i>
<i>*ā</i>	<i>*ē</i>	<i>*ō</i>

3. The entity *A. (Classical theory counted this among the vowels; here, it is placed by itself in anticipation of later discoveries.)

4. *Coefficients*, i.e. entities that by themselves sometimes establish and sometimes do not establish a syllable. There are six of these, but they can appear both as syllable-forming and as non-syllable-forming, and, in the first case, both as short and long. When we take these possibilities into account, the total becomes $3 \times 6 = 18$:

syllable-forming short:	*i	*u	*r	*l	*n	*m
syllable-forming long:	*ī	*ū	*r̄	*l̄	*n̄	*m̄
non-syllable-forming:	*i̇	*u̇	*ṙ	*l̇	*ṅ	*ṁ

A combination of a vowel and a coefficient (e.g. *ai̇*eu̇ *or* *āl̄ *ēn̄ *ōm, etc.) is called a diphthong.

Each of these symbols, then, is an abbreviated formula for a set of functions that are present under mutually exclusive conditions. Obviously, there is nothing to prevent our using these symbols in combination when we need to indicate that the related languages have two or more combined EE's, each of which corresponds to a particular symbol in the system. Indeed, we have already done so in the case of the Indo-European diphthongs, which are just such combinations of symbols. We did the same on pp. 32 and 33 in operating with IE *sp and *st, and on p. 34 in operating with Common Romance *ct. And we can combine more symbols, as a matter of course, whenever we need to do so. We need not always do so; in fact, it quite rarely happens that a whole word or parts of a word are completely identical in the different languages ("identical" here means "capable of being registered as consisting of the same Indo-European entities"). If, for example, we look at the words for 'moon' and 'month', or at the words we have collected under the rubrics *me*, *murder*, and 'winter' (p. 28), we see that they are not identical in all respects: they have been formed according to the special rules of word formation in each language. Only certain components of each word, certain EE's, remain unchanged from language to language, and these are the ones that the linguist abstracts and uses as bases for his constructions. We saw above that the cited words all include an IE *m, just as we observed (pp. 28, 32) that the words for 'tame' are identical in containing a representative of *m and a representative of *d in all the languages cited. Several of our illustrative words have been identical in respect of many other EE's – otherwise, they would not have served as evident examples for the nonspecialist. But in actual practice it is very seldom

that we can trace a whole word through the entire Indo-European family and find it remaining fully identical with itself in all the languages. It happens, however, that *brother*, *mother*, and *father* can be traced as whole words, at least if we ignore the Greek variant *phrátōr*, with its divergent vowel in the last syllable, and Lithuanian *mótina*, which has an extension containing *n*. They can therefore be formulated in their entirety as IE **bhrátēr*, **mātér*, and **patér*. In the same way the Romanist can identify each set of words on p. 34 and represent them by the formulæ **factu*, **lacte*, **nocte*. On the other hand, once we have set up our whole system of formulæ, we may often find it useful to take a word appearing in a given language – perhaps, in that particular form, only in that language – and, on the basis of our general knowledge of the element-functions, to transpose it into an Indo-European formula. Obviously, nothing stands in the way of our doing so. We can say, for example, that while all the other forms of the word *brother* represent IE **bhrátēr*, Greek *phrátōr* is to be interpreted as IE **bhrátōr*. In this way we can take any word whatsoever in any Indo-European language whatsoever and “translate” it into an Indo-European formula.

We have said that words often differ somewhat in the different Indo-European languages because they have been formed according to the special rules of word formation in each language, but this should not be taken to mean that the rules of word formation shift radically from one language to another. So long as we keep to the languages that can be most immediately accounted for by the Common Indo-European system of formulæ (Gothic, for example, rather than Modern Danish, Latin rather than French), we find, on the contrary, that word formation follows the same rules in all the languages, only that each language, for each individual word, enjoys a certain freedom of choice from among several common rules. That is to say, the system of formulæ for sets of EE-functions is not the only working tool that we can set up; in addition, we can have a system of rules of word formation that hold for all the languages being directly compared. The alternation between **ē* and **ō* that we have just now observed in the Greek word corresponding to English *brother* reflects part of just such a rule. A very important feature in the Indo-European rules of word formation consists of the so-called *vowel alternations*, according to which a certain vowel (or, as it is called, a certain alternation grade) will appear under certain grammatical conditions, depending also on relations of accent and on the alternation grade found in other syllables of the word. But these rules allow for a certain margin of variation, so that sometimes two alternation grades may appear in one and the same word. Moreover, the individual Indo-European languages

make somewhat different application of the same rules. We shall not go into these here, but merely note that the most important vowel alternation in Indo-European is between **e*, **o*, **Ø*, **ē*, and **ō*. The Greek verb ‘to fly’ (see above, p. 31) has the *e*-grade between the *p* and the *t* in the present infinitive *pétesthai*, while the perfect has the *o*-grade (together with a prefixed reduplication syllable *pe-*), so that ‘I have flown’ is *pe-pót-ēmai*. Another past tense (the aorist) of the same verb has the zero grade, that is, no vowel between the *p* and the *t*: ‘I flew’ is *e-pt-ómēn*. When the vowel **e* is the first part of a diphthong, the alternation **e* : **o* : *Ø* will of course appear as an alternation **e*₁ : **o*₁ : **i*, or **e*₂ : **ou* : **u*, or **en* : **on* : **η*, and so on. The vowel alternation that we find in English strong verbs depends on these Common Indo-European rules of word formation. Forms like *sing*, *sang*, *sung* contain, respectively, IE **em*, **on*, and **η* (**η* appears in Germanic as *un*).

From what has been said, we can conclude that *genetic relationship is a function between languages consisting in the fact that each expression element in each of the languages has function to an expression element in each of the other languages; and we have seen that each particular element-function is conditioned by the other expression elements forming the environment of the expression element concerned and by the place that it occupies in the word.*

This is what is called an *operational* definition: it is a definition that provides a criterion, a test to decide whether genetic relationship is present in a given case or not. As can be seen, the definition is based on the functions between the languages and on the functions between their expression elements. In itself, it has nothing to say about the kind of function connecting the expression elements of the different languages with the common formulæ – or about the kind of function connecting each of the languages with the system of the common formulæ. Although we have not yet closely examined the nature of this function, we shall nevertheless give it a name and call it a *continuation*; we shall say that the expression elements in each of the languages *continue* the common formulæ and that system of each language – or, in general, that each language – *continues* the aggregate system of the common formulæ. And later on we shall be able to inquire into the meaning of such statements.

When, on p. 28 we began to set up element-functions, we made the general reservation that they do not hold without restrictions. The restrictions on the domains of the element-functions must be taken into account. They do not constitute *exceptions* to the element-functions, if by exceptions we understand arbitrary and irregular violations

of a rule. They constitute *counter-examples*, i.e. strictly defined and strictly delimited regular domains for which the element-functions do not hold and which must therefore be excluded from our calculations before we establish the element-functions. The element-functions are like the Danish laws that are not promulgated for the Faeroe Islands. Once we know that fact, we can exclude the Faeroe Islands from consideration. Outside the Faeroe Islands, the law holds without exception; on the Faeroe Islands, other laws hold.

The counter-examples are never such that certain expression *elements* in a language are unaffected by the principle of the element-functions. We have already said that each expression element in a language contracts element-functions with one or more expression elements in the other languages. Thus it is always the entire system of the expression elements that is affected by the element-functions, and we never find any enclave, or any individual elements, within that system that we have to exclude from consideration. It is not merely certain elements in the one language that have function to certain elements in the other, but all to all; it is the one *linguistic structure* that has function to the other.

The counter-examples consist rather in the fact that certain signs (certain words, for example) in a language are not affected by the element-functions in the same way as other signs. We find certain kinds of signs (including certain kinds of words) whose elements do not contract the same element-functions as do the elements of other signs (other words), and we must therefore identify such special signs and exclude them from consideration when we are registering the element-functions. The counter-examples do not concern the elements as such, but only the signs composed by the elements and the functions contracted by these signs and their elements in the *linguistic usage*. Before we proceed to the counter-examples, therefore, it will be necessary to say a few words about element and sign, about the different kinds of functions between elements and between signs within one and the same language, and about linguistic structure and linguistic usage.

Linguistic Structure and Linguistic Usage

Every language appears to us first of all as a system of *signs*, that is to say, a system of expression units that have content, or meaning, attached to them. Words are signs of this sort. But parts of words can also be signs: *-s* in English is a sign of the genitive (*Jack'-s father*) and a sign of the third person singular present (*he write-s*). A word like *inac-iv-ate-s* is a sign consisting of five different smaller signs. A sign may consist of one expression element *s* with its attached content element 'genitive'; or it may be formed – both on the expression side and on the content side – by the combination of two or more elements, like the Latin sign *-ārum* in *bon-ārum mulierum* 'of the good women', which consists of four expression elements – *ā*, *r*, *u*, and *m* – and three content elements – 'genitive', 'plural', and 'feminine'.

When we look at a printed or written text, we see that it consists of signs and that these, in turn, consist of elements proceeding in a certain direction (Latin letters from left to right, Hebrew letters from right to left, Mongolian letters from top to bottom – but always in some definite direction). And when we hear a spoken text, we find that it consists of signs and that these, in turn, consist of elements proceeding in time – some come earlier and some later. The signs form *chains*, and the elements within each sign likewise form chains. We shall use the term *relation* for the function (dependence, relationship) between signs or between elements within one and the same chain: the signs, or the elements, are *related* to one another in the chain.

If we take as an example an English sign like the word *pat*, we can form other signs by inserting other elements in the place of each element within it. This is possible with regard to both the content elements and the expression elements, but we shall here confine ourselves to the latter, which are the more easily observed. Thus, we can replace each expression element in the sign *pat* with other expression elements to form such new signs as *sat, rat, tat; pet, pit, pot; pan, pad, pal*. We can set up the sign *pat* as a horizontal chain proceeding from left to right and then place under each of its elements, in a vertical column, other elements that could be inserted in its place:

<i>p</i>	<i>a</i>	<i>t</i>
<i>s</i>	<i>e</i>	<i>n</i>
<i>r</i>	<i>i</i>	<i>d</i>
<i>t</i>	<i>o</i>	<i>l</i>
.	.	.

Such vertical columns we shall call *paradigms*. Thus, a paradigm of elements will mean a class of elements that can be inserted in one and the same place in a chain: *p a t* is a chain; *p s r t* is a paradigm. The function between the members of a paradigm we shall call a *correlation*. Between *p*, *a*, and, *t* we have a relation; between *p*, *s*, *r*, and *t* we have a correlation.

Now the peculiar fact is that we cannot introduce just any elements into any paradigm. By doing so, we might (as we shall shortly be interested in observing) produce signs previously unused in English, like *miv* or *jeg*, which are still linguistically possible in the sense of being permitted by the rules governing element combinations in English, but we might also produce groups of elements, like *pgt*, or *pkt*, that would conflict with those rules. Although the elements are all right, they are combined incorrectly and the sign is impossible.

We can thus construct a paradigm consisting of elements that can be inserted only in certain places in the chain and not in others. We shall call such a paradigm a *category*. Thus, *a*, *e*, and *i* might belong to one category, and *g* and *k* to another.

When, however, we speak of restrictions on the formation of signs, we are making a very superficial observation based on our preliminary view of language as it first appears to us, as a system of signs. If we wish to penetrate more deeply and inquire what these restrictions in sign formation depend on, what structural peculiarity of language prevents a sign from having just any appearance whatsoever, we shall find the reason in the fact that there are certain rules for the construction of the syllable. There cannot be an English sign *pgt* or *pkt* for the following reason:

1. There cannot be a syllable *pgt* or *pkt* in English. If there could be, then *pgt* and *pkt* might well be English signs, just as well as English *pat*, *pet* and *pit*, or as *miv* and *jeg*, which are unused sign-possibilities in English but still sign-possibilities, precisely because they are formed in accordance with the English rules for syllable formation.

2. There cannot be a syllable in English ending in *-pgt* or *-pkt*. If there could be, then *pgt* and *pkt* might well be English signs – like *-s*, for example, which does not form a syllable by itself but which can be used as a sign in English because English syllables can end in *-s* (*cats*, *rats hats*, etc.).

3. There cannot be a syllable in English beginning in *pgt-* or *pkt-*. If there could be, then it would be possible for *pgt* and *pkt* to be English signs, since in such a case they

might always be used by being attached to the beginning of a syllable, as *-s* is always attached at the end.

4. There cannot be a syllable in English ending in *-pg* or *-pk*, or beginning in *gt-* or *kt-*. If there could be, then again it would be possible for *pgt* or *pkt* to be English signs; they would only have to be so used as always to include a syllable boundary: *apg-ta* or *ap-gta*, etc.

Thus we find that the sign-possibilities depend directly on the syllable-possibilities. The structure of language is not so ordered as to have special rules governing the shape of a *sign*. A sign may be a syllable (e.g. *pat*) or not (e.g. *-s*), and it may consist of one or more syllables (*category*, for example, consists of four). But there *are* special rules in the linguistic structure governing the shape of *syllables*, and signs must, of course, be so constructed as not to violate them.

Now what is a syllable? A syllable is not a sign (the word *category*, for example, consists of four syllables, but not of four signs). A syllable may quite accidentally coincide with a sign, since a sign may consist of one and only one syllable (as in the case of *pat*), but even then syllable and sign are not the same. The first part of *category* is the same syllable as *cat*, but not the same sign. A syllable is something different: it is a *unit of elements* resulting from the fact that certain elements contract relation with one another. The structure of the syllable depends on the relations that can be contracted by the elements – relation governed by special rules for each element or each category of elements. We said above that Eng. *a*, *e*, and *i* belong to one category and *g* and *k* to another. The category to which *a*, *e*, and *i* belong is called the category of *vowels*; the category to which *g* and *k* belong is called the category of *consonants*. Vowels each have the property of being able to form a syllable by themselves, while consonants do not. Vowels and consonants may contract relation with one another within a syllable, or combine to form a syllable. But there are certain restrictions on relations between vowel and vowel, and between consonant and consonant, within one and the same syllable. Not every element can be connected with every other but only some with some. And the rules governing relations may be such that, for example, certain consonants may be combined in a certain order but not in all conceivable orders: an English syllable may end in *-pt* but not in *-tp*.

Thus each element in a language has its place in a certain category, which is defined by certain definite relational possibilities and which excludes certain others. All such categories, together with their definitions, constitute the element-system of the

language, or what we call its *linguistic structure*. The linguistic structure determines which are possible syllables of the language and which are not – and hence, in turn, which are possible signs of the language and which are not.

It follows from the linguistic structure of English that we can form such syllables as *pat, sat, rat, pet, set, ret, pit, sit, rit, pan, san, ran, pen, sen, ren, pin, sin, rin, pad, sad, rad, ped, sed, red, pid, sid, and rid*. These, and others, can be read off from the little schema on p.33. And, in fact, all these syllables are found in English words. Even so, a syllable derivable from the linguistic structure will not always be actually used. The linguistic structure permits the formation of English syllables beginning with *spr-*, but while such possibilities as *spring* and *sprat* are actually used, others, like *spreb* and *sprust*, are not. No rules can be set up to determine the use that is made of possibilities offered by the linguistic structure; whether a given possibility is used or rejected is a matter of pure chance.

Much the same is true of signs. Since a sign, like any other chain in the language, must conform to the rules of syllable formation found in the linguistic structure, it follows that certain sign formations are excluded. But it also follows that certain others are permitted. The chains *pat, sat, rat*, etc., cited above, are possible signs since they are possible syllables, but possible syllables can be found like *pid* or *maf*, that are not signs in English. They are unused sign-possibilities, and the fact that they are not used, while certain others are, is pure accident. Moreover, we are free to use them any moment we wish, we can decide tomorrow to introduce an English word *pid* or *maf* with some meaning or another, and we can do so because they are possible syllables – the elements entering into them are given the possibility, in the linguistic structure, of occupying these places.

As we have said before, a language first appears to us as a system of signs. But we now see that in fact a language is primarily something different, namely a system of *elements* appointed to occupy certain definite places in the chain, to contract certain definite relations to the exclusion of certain others. These elements can be *used*, in conformity with the rules that govern them, to form signs. The number of elements and the relational possibilities of each element are laid down once and for all in the linguistic structure. Which of these possibilities will be exploited is determined by the *linguistic usage*.

We can conclude from this that if we wish to describe a language, the worst conceivable procedure will be the one that might appear, from, a superficial and external

point of view, to be the most natural one and the only one possible – that of beginning with an enumeration of the signs that are used in the language. Obviously, if one wishes to learn a language it is not enough to know the linguistic structure – one must also know the linguistic usage. But the usage presupposes the structure, so one's labors are considerably shortened if one begins by studying the structure. The number of expression elements in a language is quite limited: twenty-off would be common; more than fifty would be very rare. The number of syllables in a language can often be written with four digits. The number of signs, on the other hand, can go into the tens of thousands and is, in fact, by the nature of language, unlimited. A list of the signs of a language – a dictionary – is always necessarily incomplete, if only because new signs are being formed while the dictionary is being compiled and many more have been formed before the dictionary finds its last user. And yet the language in which these new signs are formed is still the same as when the dictionary was begun. A language remains the same as long as the linguistic structure is the same, and one and the same language can be subjected to different linguistic usages and put to different applications. English continues to be English even if new words are introduced into it, provided only that the words are formed according to the same rules for constructing syllables and are composed of the same elements as before.

Thus it is the linguistic structure, and that alone, that determines the identity and constancy of a language. As long as we have the same linguistic structure, we can say meaningfully that we have the same language. Without this criterion we should have to be content with observing that a language is forever changing and that each morning we wake up to a different language from the one we had the day before. Each morning we find in the newspaper some sign or other that is new in English, or at any rate new to us. Without this criterion, moreover, we should never be justified in saying that we knew a language: nobody knows and understands all the signs that have been used in English or all the signs that are being used daily by specialists in this or that field, or by people of certain regions or certain milieus.

Linguistic structure not only determines the identity of a language; it is also what essentially determines the difference between languages. A person who knows only one language or a few languages of approximately the same structure will be inclined to believe that the rules for combining elements have, so to speak, a natural inevitability. And if he is told that some combination or other is not permitted in English, he will perhaps be inclined to think that this is simply because the combination cannot come from

human lips. Actually, it is only his own habits of pronunciation that are involved. With respect to combinations of sounds, there is very little that is impossible. Our speech organs are natural Jacks-of-all-trades. We have simply got them into the habit of being able to produce only a very few of the movements that they could produce if we had trained them differently, i.e. for purposes other than those laid down by our native language. *vlk* and *krk* are impossible words in English, but perfectly good words in Czech (where they mean 'wolf' and 'neck' respectively); *lgat'* and *rvat'* would be unthinkable as English words, but are all right in Russian (with the meanings 'to lie' and 'to tear'); Georgian (a language spoken in the Caucasus) can furnish words that a native speaker of English, because of his speech habits would believe to be physically impossible – words like *vhsdšam* 'I eat', *mtha* 'hill', *mkbenare* 'biting', *dsqali* 'water'. A word like *squelch*, on the other hand, would cause a Finn or a Frenchman insuperable difficulties, if only because Finnish permits no consonant clusters at the beginning of a syllable and French no consonant clusters at the end.

The relationship between elements and signs that we have been discussing here is the real secret of the whole marvelously practical mechanism of language – indeed, one might be tempted to say, the touch of genius in the construction of language. It is always possible to form new signs simply by putting together in a new way the same familiar elements according to the same familiar rules, and both the elements and the rules are few and quickly learned. Given once and for all a handful of elements together with their rules of combination, we have an inexhaustible number of possible combinations and hence of signs. The system of elements is closed, once and for all, but the system of signs is productive; the elements constitute a closed set, but the signs an open one; within any one language the number of the elements is unalterable, but the number of signs can be augmented according to the needs and pleasure of the society or of the individual (e. g. the poet or the technical specialist) and can also be diminished as certain words drop out of use and are discarded because they are superfluous or unwanted. Being so fluid, the sign system is not merely applicable to certain situations but adaptable, without restriction, to new situations of any kind. This is why no language is bound to any particular conceptual sphere, to any particular milieu, or to any particular culture. Experts on American Indian languages have rightly emphasized that those languages would be just as well qualified as any other to express Western European culture even though, so long as they serve Indian culture, they have not yet formed signs for many of our concepts, such as those of technical and scientific origin. Let the need arise, and they would be able

to form such signs in a completely suitable way. Every language has, in addition to signs already in use, a practically inexhaustible reserve of unused possible signs.

Consequently, one can imagine a language used in many different ways. The same linguistic structure can be matched with quite different linguistic usages. One could take the whole English dictionary and replace all the words it contains with other, newly formed words that it does not contain and still preserve the English linguistic structure intact, simply by using the same elements, and only those, and combining them in syllables according to the same rules.

But while many different linguistic usages can thus be matched to one and the same linguistic structure, the converse is not true. Only one linguistic structure can be matched with a given linguistic usage. In the function between linguistic structure and linguistic usage, the structure is a *constant* and the usage a *variable*. This is what makes it reasonable to say that it is the structure and not the usage that determines the identity of a language and defines one language as opposed to another.

The linguistic structure, then, fixes the number of elements the way which each of these elements can be combined with others. That is all. All other phenomena observed in language may vary in relation to this, and they therefore constitute usage. Examples are the formation of signs and the utilization of possible signs, but other phenomena as well are variables in relation to the linguistic structure and therefore also belong to the linguistic usage. The representation of the elements (cf. p. 13) is not fixed by the linguistic structure and may vary arbitrarily with respect to it. The elements in a linguistic structure may be represented in any way whatsoever, provided only that the elements required by the structure are kept distinct. The elements may, for example, be represented graphically, with each element having its own letter. So long as the letters are distinct from one another, they may have any shape desired. We are free to choose and vary their form, color, etc. while still serving the linguistic structure. We could represent the same linguistic structure with a freshly invented writing system with a code of special conventional signs, and the language would still remain the same. Thus it is possible to represent the English expression elements in non-Latin alphabets or in Morse code and still have perfectly good English. The elements can also be represented phonetically, each element by its own sound, no matter what, so long as it is sufficiently distinct from the others. (The human ear can be trained to hear very fine nuances, so what was said above about the organs of speech applies here: the ear, too, is a Jack-of-all-trades, but the special training it has had in hearing the nuances that play a role in the native language reduces

its ability to hear the nuances that play a role in other languages, so that a person will often be inclined to declare it physically impossible to perceive the distinctions utilized in a language not his own.) It is equally immaterial how the sounds are produced – whether by phonograph, sound film, signal-whistle, or human mouth and throat. The elements of the linguistic structure can equally well be represented by other means such as flags of different designs and colors. This is the way in which ships all over the world communicate with one another in English or any other language, using an international signaling system. Individual navies have such flag codes of their own, by means of which they communicate in their native languages. The manual alphabet of deaf mutes is another special way of representing the expression elements of a language. We could go on citing such examples, and to the extent of our inventiveness, we could construct ever new means for executing one and the same linguistic structure, new means for speaking one and the same language.

The elements of the linguistic structure call to mind the entities used in algebra – like a , b , c , x , y , z – for which we can substitute different numerals in arithmetic and different numbers, with their different names, in practical reckoning. So long as we abide by the expressed conditions, we can represent the algebraic entities as we will.

To use a comparison that may perhaps be carried further, we can say that a language is organized like a game – like chess or like a card game, for example. The elements are the pieces or the cards. Different languages, like different games that differ in whole or in part. These rules state how a given element, whether piece or card, may be used and how not. To some extent, they restrict the possibilities of combination, but the number of permitted combinations or sign formations, in a language as in chess, is still enormous. The totality of the rules, stating how many pieces there are in a game and how each piece may be combined with others, can be called the structure of the game; and it differs from the linguistic usage. An account of the usage of the game would have to include information not only about the way one is *permitted* to act (this is the structure of the game), but also how people are accustomed to acting in given situations or have in fact acted up to now (this is the usage of the game) – in other words, what combinations have time-honored status under given conditions. In the same way an account of the linguistic usage would have to include information about the customary use of signs at a definite point in time in a definite milieu under given conditions. An account of the usage of the game would also have to include information about the material of which the pieces are usually made, or have been made, or are most appropriately made, and would have to

tell us how each piece or kind of piece usually looks or has looked (the knight in chess looks like a horse's head, and so on) – and in just the same way, an account of the linguistic usage would also have to include a description of the material (graphic, phonetic, and so on) of which the elements are made, and would have to tell us in detail how the individual graphs, sounds, etc., are constituted. Every usage of a game presupposes the structure of the game, but not conversely. And many different games usages can be matched to one and the same game structure. The structure of a game remains the same, whatever use is made of its rules. This is precisely the reason why the same game structure can be used over and over again for new combinations – new games – just as the same linguistic structure can be used over and over again to form new signs. And the structure of the game remains the same, even if the pieces are made of a different material or given a different shape. In this respect the structure of the game allows complete freedom, with the one restriction that pieces differing in their rules of combination must be sufficiently distinguishable. In this connexion, it may be recalled that one can play chess by telegraph, without using pieces at all but representing the elements of the game by signs in Morse code.

In using a language, people will easily come to associate with it, more or less consciously, certain notions of very different kinds. These notions may also be counted as part of the linguistic usage; in any event, they clearly do not belong to the linguistic structure. As a rule, they are notions directly associated in one way or another with the external shapes of the element-representatives – with the shapes of sounds, for example, or the letters. It will be felt that some are beautiful and others ugly, or that some are similar and others not, or that some resemble things outside language while others do not. An example of this last situation is what is customarily called *sound-symbolism*. Because words like *little*, *bit*, *kid*, or *pin* include an element phonetically represented in the linguistic usage by a certain sound (written *i*), and because in one way or another their content of meaning suggests the notion of smallness, the definite impression may be formed that there is some vague connexion between that sound and the notion of something small. It must be noted at once that these questions of similarity or dissimilarity between sounds, or between letters, or between sounds or letters on the one hand and things outside language on the other, are purely subjective, depending entirely on the point of view from which the comparison is made. Two things are never similar in every respect, but only in some respect; and given sufficient ingenuity we can always find some respect in which two things are similar, however different they may be. On my desk is an

object which, in respect of shape, resembles a dog, in respect of material, it most resembles a bunch of pipe cleaners; in respect of color, it resembles dirty snow; and so on. In the same way, with enough imagination, any two speech sounds or any two sign-meanings can be found to resemble each other.

All these notions attached to the external shapes of the element-representative may, of course, also be attached to their usual combinations as permitted by the linguistic structure – to signs, for example, combinations of signs, turns of speech, and so on. And many of these notions, because of their particularly close connexion with the material of language or because the users of a language at a given point in time and in a given milieu are for some reason or other so disposed, become collective notions which most members of a linguistics community hold more or less clearly – or can easily be led to believe that they hold. We call the totality of all such notions a *feeling for a language*, or *Sprachgefühl*. Obviously, this is a psychological phenomenon, the study of which requires the collaboration of psychologists and linguists; as yet, it is for the most part unexplored territory.

Sign formation

Since the linguistic community is free to introduce new signs and discard old ones at will, it is obvious that the Sprachgefühl, especially in so far as it is collective, will be a decisive factor in sign formation. The counterexamples to the element-functions, which we are now about to discuss, are due precisely to the fact that signs can be translated or created on the basis of a given Sprachgefühl. Since, as we have seen, such feeling for a language is unexplored territory and, by its very nature, quite intangible, we might seem to be in a poor position to make any precise delimitation of these counter-examples. If we can nevertheless do so, it is because we can be satisfied, to a large extent, with substituting for that feeling the objective factors (phonetic, linguistic, etc.) to which it is attached.

We shall now review the most important kinds of sign formation and show how they produce counter-examples to the element-functions.

II. FUNCTIONAL TRANSFORMATION OF SIGNS

I. 1. Action of elements

Within a chain of elements (a word, for example), one element may cause the appearance at another place in the same chain of a different element from the one expected. For example, corresponding to the preposition that appears as Goth. *fair-*, Ir. *air-*, Gr. *pér, péri, perí*, Lith. *per-*, OCS *prě-*, and OI *pári* (having such meanings as ‘about, through, for’ in the different languages), the Indo-European element-functions would lead us to expect Lat. *per*, with *r* according to the function given on pp. 16-17. And indeed, the word actually is *per* in Latin. Now this word can be compounded with the word *ager* ‘field, territory, territory country’ to form the verb *per-agrāre* ‘to through-country’, i.e. ‘travel through, wander through’ and the adverb *per-egrī, per-egrē* ‘abroad’ (with *e* in place of *a* under conditions that we can ignore for the moment), from which is derived *per-egrīnus* ‘foreigner, stranger’. All this is in agreement with the element-functions. But the word may also appear in Latin as *pelegrīnus*, a form that we know from inscriptions. And the various Romance forms of the same word (It. *pellegrino*, Sp. *pelegrinom* Fr. *pélerin*) presuppose the Vulgar Latin formula **pelegrinu*. This must also be the Latin form that was carried over into Germanic, as in OHG *piligrīm*, Dan. *pil(e)grim*, and Eng. *pilgrim*. But this form with *l* is not in agreement with the element-

functions. It is due to the action of one element in the word, the *r* in the syllable *-grī-*, which has caused the unexpected appearance of *l* in place of the *r* in the syllable *-re-*.

The reason for this counter-example is obviously that the same element, *r*, would have appeared in two places in the word. The counter-example has a psychological cause, namely the fact that under certain conditions it is more difficult to maintain two elements of the same kind than two different ones. Since this is a psychological matter, we can never predict whether it will or will not occur. But it has proved to be possible to establish general laws by which we can predict (1) what particular structure characterizes the words that may be replaced by such transformations, so that we can remove them from consideration when studying the element-functions, and (2) which element in a word of given structure will be affected by the change if the change takes place. Before demonstrating this, we can state in general that the replacing element introduced when a sign is so transformed will always be an element already admitted by the linguistic structure concerned. The transformation never involves an increase or a diminution in the inventory of elements in the language, and it takes place in agreement with the rules for sign formation that can be deduced from the rules of the linguistic structure governing the use of elements. Within these limits, however, there is a certain freedom with respect to the *choice* of the replacing element (provided the linguistic structure allows a choice between two or more possibilities at the given place in the syllable). The speakers choose the element that seems to them, on the basis of their feeling for the language, to lie nearest to the displaced element. It is always possible, however, to point out objective physical (most often, physiological) reason why one sound will strike speakers of the language as being especially close to another sound, so that, given a general knowledge of the language's phonetic system, we can predict with a fair amount of certainty (in many instances, with complete certainty) which element will be chosen as replacement. If we make such a preliminary calculation about *r* on the basis of a phonetic analysis of the Latin sound system (which we shall not undertake here), we arrive at the unambiguous conclusion that the only possible replacement is *l*.

As to which element in a word of given structure will be affected by the change, if the change occurs, three factors will be decisive:

1. whether the elements involved enter into a stressed or unstressed syllable;
2. whether they are found in the *initial*, *central*, or *final field* of the syllable (in a syllable like *stand*, *s* and *t* are in the initial field, *a* is in the central field, and *n* and *d* are in the final field; in a syllable like *at*, the initial field is lacking, *a* is in the central field,

and *t* is in the final; in a syllable like *ma*, *m* is in the initial field, *a* is in the central field, and the final field is lacking; in the syllable *a*, both initial and final fields are lacking);

3. whether they are found *alone* in the field (like *t* in *at*, or *m* in *ma*, or *t* and *p* in *tap*) or enter into a group (like *s* and *t*, or *n* and *d*, in *stand*), and whether they are *covered* or *uncovered* (a covered element is an element in initial field following immediately an element in final field of the preceding syllable: the *t* in *pantry* is covered, the *t* in *poetry* is not).

With the help of these definitions we can fully establish general laws stating which element in a word may be subjected to the action of another element. In what follows, we shall examine only a few examples of such general laws.

I. 1 a. *Dissimilation*

In dissimilation, speakers avoid repeating the same physiological movements in the representation of two elements. The word *pelegrīnus* is an example: replacement of *r* by *l* eliminates repetition of the tongue movement needed to pronounce *r*. Another example is Lat. *tenebrae* 'darkness', corresponding to OI *támisrā*. From the element-functions (cf. p. 15) we should expect a Latin *m* in correspondence to Old Indic *m*. The *n* that we find in Latin has been introduced by dissimilation, eliminating the repetition of lip movements used in the pronunciation of both *m* and *b*. From this last example we see that dissimilation may occur even if two elements are not completely identical, just so long as they have in common some characteristic movement of the organs of speech. In most cases of dissimilation, however, there is complete identity, so that most of the vocabulary that may be affected by dissimilation, and that must therefore be removed from consideration when we are studying the element-functions, consists of words in which the same element appears twice, with one or more elements in between.

The fact that in *peregrīnus* it is the first, and not the second, *r* that may be affected by dissimilation is due to the general law that an element in a group in a stressed syllable dissimilates an element standing alone between two vowels – never the other way round. Thus it is inconceivable that a form *pelegrīnus* would arise from dissimilation; if *peregrīnus* is subjected to dissimilation, *pelegrīnus* is the only possible result. By virtue of the same general law, *cérébral* 'cerebral' is dissimilated in popular French to *célébral*, *contrario* 'contrary' in Galician (spoken in northwestern Spain) to *contralio*, *imperatrice(m)* 'empress' in Portuguese to *empānatriz*, and so on.

The fact that in a case like *temebrae* it is the *m* and not the *b* that may be (and in this example actually was) affected by dissimilation is due to the general law that, of two elements having the same definition according to the criteria on pp. 47-48, it is always the first, and never the second, that is dissimilated. In *temebrae*, *m* and *b* have the same definition in this respect: both are in unstressed syllable (the stress is on the first syllable of the word), both are in the initial field of their syllable, and both are uncovered. The syllable with *m* comes before the syllable with *b*, so it is the *m* that must give way. By virtue of the same law, Fr. *militaire* is dissimilated in popular speech in *mélitaire*: both *i*'s are in unstressed syllable and in the central field – hence the first *i* is dissimilated, not the second.

If Fr. *barbier*, derived from *barbe* ‘beard’, can appear in German as *Balbier* and in Danish as *balbér*, but not the other way round as *barbiel*, *barbél*, this is due to the general law that an element in the final field of the stressed syllable dissimilates an element in the final field of the unstressed syllable, and not the other way round. This is why we also find Lat. *arbor* ‘tree’, Fr. *arbre*, appearing in Spanish as *árbol*, and Lat. *marmor* ‘marble’ appearing in Spanish as *mármol* and in German and older Danish as *Marmel*.

Another such general law states that a covered element dissimilates an uncovered one, never vice versa. This explains the change of the Latin numeral *quīnque* ‘five’ to *cinque*, which is the basis for Fr. *cinq*, It. *cinque*, etc. The Italian word *tatufolo* was carried over into German as *Tartuffel*, but in accordance with this general law of dissimilation, was changed in both German and Danish to *Kartoffel*, ‘potato’. Dissimilations like *quince* or *tarkoffel* would be inconceivable.

I. 1 b. *Metathesis*

Metathesis (in the restricted sense in which we shall use the word here) is a transfer of an element from one syllable to another. It is found in two forms:

1. Transfer of one element. This always consists in the transfer of an element in a group from an unstressed to a stressed syllable – never otherwise. By this form of metathesis Italian has acquired the word *cocodrillo* in place of Lat. *crocodilus*. The form with metathesis evidently existed in Late Latin and was carried over into Middle High German, where it is *kokodrille*.

2. Interchange of two or more elements. In the simplest and most frequent cases, two elements in neighboring syllables are interchanged. This always takes place in such a way that elements not appearing in the order of expiration are transposed so that they do. (Order of expiration is the order of movements of speech organs from the interior to the exterior – from throat to lips.) We have discussed above (p. 19) the Latin word *specere*, a word also known from Old Indic, where it appears as *spásati*. These forms establish the Indo-European formula for the root as **spek₁-*. But in Greek the word meaning ‘to look, view, examine’ is *sképtesthai* (from which are derived the well-known word *sképsis* ‘examination, speculation’ and *skeptikós* ‘one who speculates or reflects’, carried over into Eng. *skeptic*). This is due to a metathesis by which the *p-k* sequence has been changed to *k-p* because the *k*-sound is formed farther back in the mouth than the *p*-sound: the *p* sound is a labial stop, while the *k*-sound is produced by an occlusion between the tongue and the palate.

Other kinds of metathesis are governed by other laws, which we shall not discuss here.

I. 1 c. *Haplology*

This term refers to the omission of a syllable on account of its similarity (most often an identity) to a neighboring syllable: *tragicomic* for *tragico-comic*, *mineralogy* for *mineralo-logy*. Almost all the general laws that we know governing the action of elements were discovered by the French scholar Maurice Grammont. Some of these laws may well need closer scrutiny, but there is no doubt that general laws can be established. Apparent exceptions can be explained by the fact that a word will often have been subjected to several successive transformations, so that a given form is not always to be explained merely by the mechanical applications of some single law. Moreover, Grammont foresaw the possibility that the blind action of a law may lead to results conflicting with the linguistic structure by producing an inadmissible type of syllable. In such a case, he calculated that the tendency to dissimilation or metathesis, if sufficiently strong, will have its effect under a set of laws opposed to those discussed above; but even then, there will be no suspension of law, since such cases can be clearly circumscribed and defined.

I. 2. Analogical formation

By analogical formation we mean an action of signs. When there is a paradigm of signs (of words, for example), one member of the paradigm may cause the appearance at another place in the paradigm of a different sign from the one expected. There are three main types of analogical formation.

I. 2 a. *Leveling*

Leveling occurs when two members in a paradigm are brought closer together in form. In the nominative singular, the second personal pronoun of Indo-European appears as both **tu* and **tū*, as attested by Goth. *þu*, Ir. *tú*, Lith. *tù*, OCS *ty*, etc.

The other cases are formed on a stem consisting of **t* or **t_u* followed by a vowel. In the accusative this vowel is **e* or **ē*, so that we start with the forms **te* or **tē* – found, for example, in Lat. *tē* and ON *þi-k* (the Gmc. *-k* corresponds to Gr. *-ge*, a particle that can be added to pronouns; cf. Goth. *mi-k*, p. 14) – and the forms **t_ue* or **t_uē* – found at the base of such forms as Hitt. *twel* ‘thee, thine’, *twetaz* ‘of thee, from thee’, or OI *tvam* ‘thou’. In Greek, some dialects have the accusative form **te* – as in Doric *té* – and others have **t_ue* – as in Attic-Ionic *sé* (where *s-* = IE **t_u-*); other cases of the same word are formed on **t* and **t_u-* with the same dialectal distribution (dative: Doric *toí*, Attic-Ionic *soí*; cf. OI *tve*). All this is in accord with the element-functions, and our paradigms are what we should expect. –

	ON	Lat.	Dor.
Nominative	<i>þú</i>	<i>tū</i>	<i>tú</i>
Accusative	<i>þi-k</i>	<i>tē</i>	<i>té</i>

In Gothic, however, the accusative is not *þi-k*, as it is in Old Norse and as we should expect from Indo-European **te-ge*, but *þu-k*. A leveling has taken place, bringing the two members of the paradigm closer together in form and thereby producing a counter-example to the element-functions: a Gmc. *u* cannot be an IE **e* (or **ē*). In Hittite, we find something even stranger: the nominative is *zik* (= IE **te-ge*) and the accusative *tuk* (= IE **tu-ge*). This paradigm is the result of two contradictory levelings, each striving to bring the two members of the paradigm closer together in form, which paradoxically enough have led to a compromise no better than the point of departure. Greek also shows a leveling in this paradigm, but in another way: the Attic-Ionic nominative is not *tú* like the

Doric, but *sú*, showing analogical influence from those cases in which the word begins with *s-* (= IE **t_u*).

Other instances of levelings that produce counter-examples to the element-functions can be found in the numerals. Indo-European ‘four’ and ‘five’ are **k₃et_uōres* (cf. Lat. *quattuor*, Lith. *keturi*, OI *catvārah*) and **penk₃e* (cf. Lith. *penkì*, OI *pāñca*). They begin with two quite different consonants, to which the word-initial element-functions will assign quite different correspondences in Germanic. IE **k₃-* appears in Germanic as *hw-* as shown in the interrogative pronoun: Goth. *hwas*, genitive *hwis*; Lat. *quis*, feminine *quae*; Lith. *kàs*; OI *kaḥ*, etc. IE **p-* appears in Germanic as *f-*, as we know from *father* and similar examples. Nevertheless, not only the numeral *five*, but also the numeral *four* begins with an *f-* in Germanic: Goth. *fidwor* ‘four’, *fimf* ‘five’. (The final *f* in *fimf* corresponding to IE **k₃* is in accordance with an element-function: **k₃* has Germanic correspondence *f* if there is a preceding **p* or **_u* in the same word – cf. Goth. *wulfs* ‘wolf’, Lith. *vil̃kas*, OCS *vl̃kь*, OI *vṛkaḥ*, IE **u̯_lk₃os*. We may also note that the first *qu* in Lat. *quīnque* corresponding to IE **p* is in accordance with an element-function: **p* has Latin correspondence *qu* or *c*, depending on the following vowel, if there is a following **k₃* in the same word – cf. *coquere* ‘to cook’, OCS *pek_o*, OI *pácati*.) The reason for *f-* in Goth. *fidwor*, Eng. *four*, instead of the expected Goth. *hw-*, Eng. *wh-*, is an analogical leveling that has brought two members of the paradigm of numerals closer together in form.

In general, levelings are common in the paradigm of numerals. The numeral *nine* begins with **n-* (cf., for example, Lat. *novem*, OI *nāva*), while the numeral *ten* begins with **d-* (see p.19), but in the Baltic and Slavic languages both numerals begin with **d-* (Lith. *devynì*, *dėšimt*, OCS *devětь*, *desětь*): the **d-* in ‘nine’ has been introduced by leveling. According to the element-functions, we should expect Lat. *quīnque* (with the dissimilation discussed on p. 49) to end in *-e* in Spanish. In fact, it ends in *-o* (*cinco*), under the analogical influence of *cuatro* ‘four’. Similarly, Italian has final *-i* in *dieci* ‘ten’ (Lat. *decem*) by leveling with *venti* ‘twenty’ (Lat. *uīgintī*).

Paradigms like those we have been considering – the case paradigm of a pronoun and the paradigm of numerals – are given in the linguistic structure itself and constitute absolutely closed series with a limited number of members. (Even though the numbers constitute an infinite series, the series of number *words* is finite and, indeed, always fairly short, with most numbers being designated by a combination of two or more words.) But in linguistic usage we are faced with many paradigms having an unlimited number of members, since, as we have seen, the sign system is not closed, but productive. Here,

where Sprachgefühl will be uncertain in its arbitrary selection of the paradigm to which a given sign shall be allotted, it will play fast and free, haphazardly, and on the basis of very loose associations. If the Indo-European word for ‘eye’ begins with *o- (cf., for example, Lat. *oculus*, OCS *oko*) and the Indo-European word for ‘ear’ begins with *au- (cf., Lat. *auris*, Lith. *ausis*), while both words begin with *au in Germanic (Goth. *augo*, *auso*, Dan. *øje*, *øre*), the state of affairs in Germanic will be due to a leveling, caused by the fact that linguistic feeling readily tends to locate the two words in the same paradigm. Ger. *Sauerkraut* has been taken over into French as *choucroute* because Sprachgefühl has located it in the same paradigm as *chou* ‘cabbage’ on the basis of similarity in meaning and has thus produced a leveling that brings the two members of the paradigm closer together in form. In the same way, Fr. *mousqueton* ‘carbine’, a derivative of *mousquet* ‘musket’, was carried over into German in the form *Muskedonner* under the influence of *Donner* ‘thunder’, and into Norwegian and Swedish in the form *muskedunder* under the influence of the word *dunder* ‘din, uproar’. This kind of leveling, which depends on loose associations in the Sprachgefühl and arbitrary linkage of two signs in a single paradigm, is called *popular etymology*.

Levelings are the only analogical formations that commonly produce true counter-examples to the element-functions. For this reason, we shall consider only very briefly the two other kinds of analogical formation.

I. 2 b. *Contamination*

By contamination is meant the blending of two members of a paradigm into one, as *irrespective* and *regardless* into *irregardless*, or *snort* and *chuckle* into Lewis Carroll’s *chortle*, or *breakfast* and *lunch* into *brunch*.

I. 2 c. *Proportional formation*

In proportional formation a sign is transferred to a different paradigm from the one to which it previously belonged. The strong and weak verbs of English constitute two different paradigms, but many formerly strong verbs have been transferred by proportional formation to the weak paradigm, producing new forms like *helped* for older *holp*. The different declensions and grammatical genders likewise constitute different paradigms. In German, the word *Nacht* ‘night’ belongs to a paradigm consisting of

feminine nouns that take umlaut in the plural (*Nächte*), while *Tag* ‘day’ belongs to a paradigm consisting of masculine nouns that have “strong” declension (genitive *Tags*) and do not take umlaut in the plural (*Tage*). The genitive of *Nacht* in the definite form is *der Nacht*, but still ‘at night’ can be *des Nachts*, just as ‘in the daytime’ is *des Tages*: a proportional formation has brought *Nacht* into the same paradigm as *Tag*.

We call these formations proportional because they presuppose the construction of a proportion or equation of the type $a : b = c : x$, with the solution $x = b'$; for example, *hinder : hindered = help : x*, therefore $x = helped$. Contamination, on the other hand, follows the formula $a \times b = c$, thus, *breakfast \times lunch = brunch*.

Since the sign system is subject to the arbitrariness of linguistic usage, and since Sprachgefühl is unpredictable, it is clear that general laws cannot be established in this area as they can for the action of elements. But perhaps this is only because the linguistic content has not yet been adequately analyzed from a psychological point of view. If the actions of elements have been so thoroughly explored, it is precisely because we are able to analyze the sounds of language and because we know something about the psychology of sound. Perhaps in the future we shall have more solid bases than at present for the study of analogical formations. Meanwhile, our experience with the genetic study of an enormous amount of linguistic material has shown that analogical formations in no way prevent us from getting to know the element-functions and, in general, cause no practical difficulties in our investigations. We have only to be aware of the fact that the principle of analogical formation provides a reason for setting aside certain counter-examples with which we may have to deal.

I. 3. Abbreviation of signs

A sign may undergo abbreviation when transferred in certain ways to a different category from the one to which it previously belonged. The most important instances are the following:

1. A compound word becomes simple. Thus, compounds originally meaning ‘loaf-ward’ and, probably, ‘loaf-kneader’ became simplified in English to result in present-day *lord* and *lady*.
2. A derivative word becomes a non-derivative. Thus, Ger. *fressen* ‘to eat (of animals), to cram’ has replaced *ver-essen* (Goth, *fra-itan*), and Ger. *gleich* ‘equal, like’

has replaced *ge-leich* (Goth, *ga-leiks*; cf. also, without the derivational prefix, Dan. *lig*, Eng. *like*, and Lith. *lygus* ‘straight, smooth, like’).

3. A sign is transferred to another grammatical category and shortened in the process. Thus, a substantive transmuted into a preposition is often abbreviated. Lat. *casa* ‘hut’ (It-*casa* ‘house’) appears in Old French as *chiese*, in accordance with the element-functions involved. But the phrase *in casa* followed by a genitive, meaning ‘in so-and-so’s house’, was transferred to the category of prepositions and took on the meaning ‘at (so-and-so’s)’. In consequence, the final vowel dropped, and the result was OF *en chies*, Modern Fr. *chez* (without *en*). The corresponding Danish preposition *hos* bears the same relation noun *hus* ‘house’; cf. also Dan. *til*, Eng. *till*, and Ger. *Ziel* ‘goal, target, term’.

Another instance of the same kind is the transfer of a verb to the category of auxiliary verbs. A Latin *b* between two vowels appears as *v* in French: Lat. *faba* ‘bean’, Fr. *fève*; Lat. *dēbēre* ‘owe, ought’, Fr. *devoir*; Lat. *habēre* ‘have’, Fr. *avoir*. But corresponding to Lat. *habeō* ‘I have’ we find Fr. *ai* with zero instead of the expected *v*. The reason is that the verb has become an auxiliary and has consequently been abbreviated. We should, then, really expect to find *j’ai* used as an auxiliary (as in *j’ai mange* ‘I have eaten’) and another form, containing *v*, used as a full verb. The fact that we find *j’ai* in this function also (*j’ai un ami* ‘I have a friend’) is due to a generalizing analogical formation.

The categories we have considered so far – compound, derivative, substantive and preposition, full verb and auxiliary verb – are given in the linguistic structure itself. But categories may also be established by the linguistic usage. Such category formation, however, is very easily delimited: a category may be established consisting of signs that are in especially frequent use and are consequently abbreviated.

4. A sign is abbreviated because of frequent use. We can begin by discussing some borderline cases which may also involve transfer to a category given in the linguistic structure. In Latin, the imperative of consonant-stem (third-conjugation) verbs ends in *-e* (= IE **-e*): *scrīb-e* ‘write!’ from the verb *scrībere*. But the five verbs *dīcere* ‘say’, *dūcere* ‘lead’, *facere* ‘do’, *ferre* ‘carry’, and *emere* ‘buy’ have imperative forms *dīc*, *dūc*, *fac*, *fer*, and *em*, without *-e*. (The form *em* has the special meaning ‘there! see!’ while ‘buy!’ is *em-e*. The verb is found with the meaning ‘take’ in other Indo-European Languages, e.g. Lith. *im̃ti*, OCS *ęti*.) Here we might say that these imperatives have been transferred to the category of interjections, which is surely to be found in the linguistic structure. We find the same situation when an answer-word like *yes* or *no* appears, as it often does, in a

form conflicting with the element-functions. Answer-words also constitute a category, which is defined by its special relation to the environments and which is therefore also to be found in the linguistic structure. Here also belong forms of address, which we often find quite radically abbreviated: Fr. *monsieur* is pronounced *m'sieu*, Eng. *madam* collapses to *mam* or merely *m*, especially when combined with an answer-word (*yes'm*). In the same way, Russ. *gosudár'*, first shortened to *súdar'*, was further reduced to *s* alone when combined with an answer-word (*da-s* 'yes, sir'). In this way doublets often arise, one of which retains its full lexical meaning while the other has been reduced to a more or less empty formula of politeness. Words meaning 'you' also frequently turn out to be abbreviations of longer complexes: Sp. *Usted* is a compression of *Vuestra merced* 'Your Grace', and Lith. *Támsta* of *tàvo mýlista* 'Thy Grace'.

But even if there is a clear connexion here with categories anchored in the linguistic structure, we still must recognize in the linguistic usage the existence of a category of signs that are used with special frequency and therefore abbreviated. Thus, not only is *automobile vehicle*, Fr. *voiture automobile*, abbreviated to *automobile*, but the latter is further curtailed – in French, English, German, and several other languages, to *auto*, and in Danish to *bil*. In the same way, *airplane* is shortened to *plane*, and Dan. *flyvemaskine* 'flying machine' to *fly*. Abbreviations of this sort are especially common in professional jargons, because certain signs are used with high frequency by profession people in a particular field (though they may be strange or completely unknown to others). English students' vocabulary provides examples like *lab* (also found in German and Danish) for *laboratory*, *exam* for *examination*, *prep* for *preparation*, and *prof* for *professor*. As a rule, in these cases, too, the shorter or exists with the longer and has no different meaning; it is not really a new sign, but a sign-variant, quite on a line with abbreviations in the written language. Moreover, such formation, unlike many others, are produced quite consciously, by a voluntary interference in the language. Some linguistic communities are more inclined to introduce them than others, and this often depends on the general structure of signs in the language. English, which already has so many monosyllabic words that it seems natural to introduce more, favors this kind of abbreviation. Danish is considerably more conservative in this respect, a fact undoubtedly connected with the structure of the language, according to which abbreviation often involves a change in pronunciation. Despite its usefulness, *fly* is a bit inconvenient as an abbreviation of *flyvemaskine* because it cannot be pronounced completely like the first syllable of the longer word: Danish linguistic structure requires the addition of a glottal

catch to the monosyllable *fly* that the unabbreviated word does not contain. Norwegian and Swedish, which do not have this difficulty, adopt such abbreviations much more easily. In recent times, as is well known, German, Russian, and English have created a large number of abbreviation-names like *CDU* for *Christliche Demokratische Union*, *DDP* for *Deutsche Demokratische Partei*, *Komintérn* for *Kommunističeskij Internationál*, *Komsomól* for *Kommunističeskij Sojúz Moloděži*, *UN* for *United Nations*. In Danish, abbreviations composed of initial syllables of words are not common (although we do find, in spoken as well as written Danish, such abbreviations of academic degrees as *stud*, *mag.*, *mag. art.*, etc.). Danish prefers abbreviations composed of initial letters – which are also common in German and Russian. Where these produce possible syllables, some languages have a tendency to pronounce them so as, for example, *URSS* may be pronounced in French.

III. INTRODUCTION OF NEW SIGNS

II. 1. Borrowings

“Borrowing” (an established, but somewhat misleading term) means the transfer of a sign from one language to another. Most frequently, words or stems of words are so transferred, but other kinds of signs may also be borrowed, like the English prefix *inter-*, which comes from Latin but which can be added to genuine English words. Borrowed words are called *loan words*. Loan words whose external forms produce an alien or incongruous effect – by reason, say, of an unusual syllable-onset (*tsetse*) or special type of accentuation – are also called *foreign words*. What is characteristic of foreign words is that they do not follow the laws of the native linguistic structure, but of some foreign structure. They are like minorities living under their own special laws. Consequently, they are often subject to transformation by unlearned speakers following their native Sprachgefühl, or to popular etymology (*very coarse veins* for *varicose veins*). As for loan words that are not foreign words, the man in the street has no suspicion of their origin and will be astounded when told that words like *skirt* (from Scandinavian), *street* (from Latin), or *chair* (from Old French) are loan words.

When a word has been borrowed from a known source, it is, of course, easy enough to demonstrate its identity, but the element-functions will not ordinarily hold. We have already had occasion to refer to words that have been carried over in this way from one language to another: Greek and Latin borrowings in Germanic, like *fraternize* (p. 11),

paternity (p.13), *pecuniary* (p. 18), *inspect* (p. 19), *pilgrim* (p. 46), *crocodile* (p. 50), *skeptic* (p. 50), and others; French borrowings in Germanic, like *barber* (p. 49) and *Muskedonner* (p. 54); Italian borrowings in French like *choucroute* (p. 54). We can see that while the element-functions have word-initial Gmc. *f-* corresponding to Lat. *p-* (*father, fee*), the Latin borrowings with initial *p-* appear in Germanic with *p-* (*paternity, pecuniary*). In this case, and in the vast majority of cases, there is an identity between the letter or sound in the lending and in the borrowing language. To this need only be added that loan words which do not continue to be foreign words are adapted not only to the native linguistic structure, but also to the native linguistic usage; such adaptation is normally the first step toward the assimilation of a loan word into the borrowing language. Thus, for example, we pronounce *psychology* without the *p* (adaptation to English linguistic structure) and a French loan word like *faux* (in *faux pas*) with the sounds we use in pronouncing English *foe* (adaptation to English linguistic usage).

There are, however, instances – although rather less frequent – when the loan word is transformed at the time of borrowing so as to accord in whole or, usually, only in part with the element-functions between the lending and the borrowing languages (provided, of course, they are genetically related). If this is to happen, the linguistic community must contain enough bilingual speakers whose Sprachgefühl makes them conscious of the regular element-functions between the two languages. Thus, there has been a consciousness in Denmark (undoubtedly influenced by the Low German borrowings in Danish) that Dan. *sk* corresponds to Ger. *sch* (pronounced [š]), as illustrated by such pairs as *skabe* – *schaffen*, *skade* – *schaden*, and *ske* – *Schuh*. Consequently, the French word *jus* ‘juice’, which is pronounced [žy] and was borrowed into German with the pronunciation [šy] by adaptation to German linguistics usage, appears in Danish, which has borrowed it from the German, as *sky* ‘gravy’.

The speakers of a language, however, are rarely such competent linguists as to carry off a “fraud” like this successfully. Sprachgefühl will normally be concerned only with certain differences between languages that are particularly striking to the ear. In practice, then, it is almost always possible to detect from a word’s external form whether it is a loan word. Indeed, the external form of the word is the linguist’s only criterion. If a word can be shown to be connected with a word in another language but fails to conform to the element-functions that have been established on the basis of other examples and cannot be explained in any other way (by the action of elements, by analogical formation, or by the introduction of a new sign in some different fashion), it must be concluded that

the word has been borrowed. Borrowing is thus an explanation that the linguist always has in reserve when other explanations fail. This does not mean, however, that we have created a convenient lumber room where we can dispose of obstinate cases by simply calling them “borrowings.” As we have seen, borrowing is something quite precise and well defined, and the assumption of borrowing is logically compelling when other explanations are impossible. In all those instances where we can trace in detail the historical relationships and cultural influences that must be presupposed for the borrowing to have taken place, we can show that, whenever we find a word not genetically related to words in the genetically related languages and not explainable by the action of elements, by analogical formation, or by the introduction of a new sign in some different fashion, then the word *was*, indeed, introduced from outside at some determinable time and by some determinable route. There are, to be sure, many other cases where the external historical relationships and cultural currents are not so clearly known as to permit any such demonstration. But in such instances we generalize, just as science always does in like circumstances: from our knowledge of the fully elucidated cases we conclude hypothetically that the less well elucidated cases are of the same kind, provided there is nothing to contradict the assumption. In practice, then, the external shape of words is our only decisive criterion in establishing the fact of a borrowing, and by this purely linguistic means we are able to make inferences about prehistoric contacts of peoples and cultural currents that would otherwise remain completely unknown because no other sources of information are available.

If we leave out of account those cases – especially from more recent times – where the external historical circumstances are clearly known in full detail, so that the very introduction of the borrowing can be dated and investigated, there are no other general guidelines for the establishment of borrowing than internal, linguistic ones. In principle, there are no limits to what can be borrowed from one language into another. Often a sign will naturally be borrowed because it accompanies the thing it designates: an article of export carries its designation with it, as when Lat. *uīnum* ‘wine’ spread as a loan word over almost all the rest of Europe. But this is only a special case. A word is not borrowed merely because the thing it designates has no native designation. People borrow out of a desire to imitate, and borrowings are subject to all the caprices of fashion. People borrow because the foreign word is finer or more beautiful or more amusing or more interesting, and nobody can predict what people will consider fine, beautiful, amusing, or interesting. Bilingual populations may also, of course, introduce borrowings into a language simply

because they have a hard time keeping the sign systems of the two languages apart. Examples of this can be found in all border regions. But this again is a special case. It is impossible before the event to say anything at all about what can and cannot be borrowed. The loan words in a language can rise to a formidable number. Albanian has been reported by some investigators to have only some six hundred words that are not borrowed, with all the rest coming from Latin, Romance languages, Slavic languages, Greek, and Turkish. Even a language like Greek, which is considered one of the purest Indo-European languages and which plays a greater role than any other in comparative Indo-European studies, contains only a relatively small number of words that can be genetically accounted for on the basis of Indo-European. Presumably, then, most Greek words are borrowings from other languages, chiefly, perhaps, from non-Indo-European languages. Just as we could observe above that some linguistic communities are more inclined than others to admit abbreviations of signs, we may say the same with respect to borrowings, without being able to offer any explanation for this difference. There are languages to which the many international words of Græco-Latin origin that abound in most modern European languages have a very hard time gaining admission, because the linguistic community prefers to use native lexical material to designate the things concerned. Thus *university* in Icelandic is *háskóli* ‘high school’, in Finnish *yliopisto* ‘institute of higher learning’, and in Hungarian *egyetem*, an artificial derivative from the numeral *egy* (pronounced [ɛd’]) ‘one’, in imitation of the Latin *ūniversitās*, a derivative from the numeral *ūnus*; while other European languages use the international word that appears in English as *university*. Icelandic is particularly famous (or, if preferred, notorious) for its stubborn resistance to linguistic imports.

We cannot predict, then, what may be made an object of borrowing, but must allow for the possibility of anything’s being borrowed. Nor can we know beforehand whether a language be resistant to borrowing, or to what degree. Nor can we establish general laws about the direction in which borrowings will move. They may pass from culturally inferior to culturally superior populations or vice versa, from languages with few speakers to languages with many or vice versa, from the language of subject peoples to the language of their rulers or vice versa, from dialect to standard language or from standard language to dialect. Wherever two linguistic communities are in contact, the possibility of borrowing must be taken into account.

In establishing the fact of borrowing, then, the linguist is generally reduced to using purely linguistic criteria, of which the most important is the fact that the words involved constitute counter-examples to the element-functions. We have seen above that Celtic corresponds with zero to IE **p*, Lat. *p*, and that, for example, the Irish correspondent to Eng. *fish*, Lat. *piscis*, is *iasc* (p. 18). Now when we discover that the Welsh word for fish is *pysg*, we must conclude that it is a borrowing from Latin. In this case, as it happens, the assumption can be supported on many sides, because we know that in general there was a considerable influence of Latin culture in Wales. It often occurs at a given period borrowings between two languages move predominantly in one direction. A general movement of loan words that played a great role in Europe proceeds from the Mediterranean countries, especially from the Semitic languages, and passes through Greek and on to Latin, from Latin to French and finally on to Germanic. The English word *sack* has come to us by this long route. Danish, as we have seen, acquired a large number of borrowings, partly from French and partly from Scandinavian. Over a long span of time, French has borrowed extensively from Latin. Vilhelm Thomsen showed (particularly in his dissertation of 1869) that Finnish – which we have seen to be quite conservative in historical times with respect to admitting foreign lexical material – must in pre-historic times have accepted a very considerable number of loan words, especially from the Germanic languages. Some of these borrowings are closer in form to the Common Germanic word formula than anything in the recorded Germanic languages, so they play a special role in comparative Germanic linguistics.

II. 2. Tabu

The word *tabu* is Polynesian and refers to the situation in which a word or name can be used in a community only under special conditions, whether only by certain persons or only in certain circumstances. This phenomenon was first observed among primitive peoples, whose religious notions often cause it to be very widespread. Recent investigation, however, has shown that tabu is known in all societies and at all times, and that it plays a quite considerable role in our own languages. The most obvious example is the fact that it is considered improper to name by name phenomena connected with sexual life or the digestive process. This tabu is not, as one might perhaps think, common to all mankind. It is linked to certain societies and is unknown in many societies outside modern Europe; it did not exist in ancient Greece. But other areas of life – the religious

area, for example – may also be surrounded by such linguistic mystery: one must not take God's name in vain. Consequently, if a person has to mention such things, he must either use allusions and circumlocutions or else simply change the words that designate them. Often, as is well known, a foreign word is used: it is not the thing itself that is under tabu, but the sign, and if one uses a foreign sign, the odium disappears. Or another word is arbitrarily chosen which actually means something quite different but whose external form resembles that of the tabu word enough to call it to mind (*sewer-pipe* for *suicide*). Or, finally – and this will be of special interest to us here – the word is arbitrarily transformed. This can be done in many different ways: by abbreviation (for example *TB* may be used instead of *tuberculosis* to avoid invoking the dreaded disease by using its real name), or by transportation of elements (Dan. *pokker* 'the deuce, the devil', if so derived from *kopper* 'smallpox', can be explained only by tabu, and not as a metathesis, since it runs counter to the general law governing metathesis; see p. 50), or by substitutions of different elements at certain places in the word. One may perhaps, for the purpose of giving an example, be allowed to observe that it can be regarded as permissible in Danish to refer to a little natural business by the verb *nisse*. This word uses *n* to replace the *t* of another word, which is under a fairly strong tabu, although the word with *t*, in turn, was originally introduced as an arbitrary substitute for a third word, which (like the corresponding English word) is under even stronger tabu. In this way, tabu produces counter-examples to the element-functions. There can often be mutual reinforcement by tabu and the action of elements. The transformation of Lat. *meretrīx* 'harlot' to *meletrīx* (**meltrīce* is presupposed by older Fr. *meautris*, Provençal *meltriz*) can be explained by a dissimilation like that found in *pelegrīnus*; but the form *menetrīx*, in which the word also appears, cannot be so explained within the framework of the Latin sound system and must be owing to tabu.

II. 3. Neologism

Finally, we must observe that it is, of course, also possible to introduce into a language completely new signs which are neither transformations nor derivations based on other signs but which follow the rules of syllable formation in the linguistic structure. Here again, different linguistic communities at different periods behave quite differently. Most Western European languages, as compared with other language areas, show surprisingly little inclination to accept such neologisms. Rather, our languages betray an

aversion to them, if we leave out of account onomatopoetic formations and the special license sometimes granted to creative writers. On the other hand, this need not always be so. To take a particularly striking example from recent times, several thousand new words were arbitrarily introduced into Estonian by government decree in this century, words that were truly created, so to speak, out of nothing, by putting Estonian expression elements together in new ways according to Estonian rules of syllable formation. Such neologisms, of course, can never be more than apparent counter-examples to the element-functions. A newly created word may perplex the investigator of genetic relationship only if it is fairly close in form to what he might be led by the element-functions to expect to find in the language. The linguist must therefore also be on the watch for this possibility. Moreover, he must keep in mind that if relatively few words in a language (Like Greek, for example) can be explained genetically, this need not always be due to borrowing but may also be due to neologistic creation.

Language Families

The method we have outlined has made possible the recognition of a large number of language families all over the world. As we have already seen, however, there are families of different degrees, so in what follows we shall reserve the term *family* to designate a large family, like Indo-European. Such a family may then be divided into *groups*, *subgroups*, *branches*, and so on. We shall now proceed to a summary review of the most important language families that have so far been established.

I. The *Indo-European* family (see p.10)

1. The *Germanic* group

A. The *East Germanic* subgroup

- a) The *Gothic* branch, which included several languages, only one of which is at all well recorded (see p. 11).
- b) The *Scandinavian* (or *Norse*) branch. The oldest stage is *Old Norse*, which, as represented in runic inscriptions, is altogether the oldest and most archaic of all the Germanic languages. The younger stage divides into two subbranches: *West Scandinavian* (*Icelandic*, *Faeroese*, *Norwegian*) and *East Scandinavian* (*Swedish*, *Danish*). From 1530 to 1814, Danish was also the standard and literary language of Norway. The Norwegian “riksmål” or “bokmål,” is a Danish slightly colored with Norwegian; the “landsmål,” a standard language formed on the basis of the West Norwegian dialect of Søndmøre, has been given equal official status with the “riksmål” since 1892 and provides the foundation for “nynorsk.” In 1938 a third standard language, “samnorsk,” was introduced as a compromise.

B. The *West Germanic* subgroup

- a) The *Anglo-Frisian* branch (*English*, introduced into England by invaders from the Continent and diffused throughout the British Empire and the United States; *Frisian*, spoken along the west coast of North Holland and Germany).

- b) The *Dutch-German* branch (*Dutch*, with two written languages: Dutch-Flemish, in Holland, Belgium, and colonies, and Afrikaans, in South Africa; *Low German* and *High German* – see p. 24).

Within Indo-European, languages of the Germanic group presently count the largest number of speakers – a good 400 million, of whom by far the most speak English, the world's most widespread language after Chinese.

2. The *Celtic* group

A. The *Gaelic* subgroup (*Irish*; *Scottish Gaelic*, introduced into Scotland by Irish immigrants in the fifth century; *Manx*, now extinct, once spoken on the Isle of Man).

B. The *Brythonic* subgroup

a) The *Continental* branch (*Gaulish*, attested by inscriptions, extinct after the conquest of Gaul and the creation of the Roman Empire).

b) The *Insular* branch (*Welsh*, in Wales; *Cornish*, in Cornwall, extinct; *Breton*, introduced into Brittany from Britain in the fifth and sixth centuries, with four different written languages).

3. The *Italic* group

A. The *East Italic* subgroup, considerably more widespread than West Italic at the beginning of historical times.

a) The *Umbro-Samnite* branch (*Umbrian*, in the Italian region of Umbria; *Samnite*, in southern Italy, with the official language called *Oscan*).

b) The *Lepontic* branch (in the northwestern Alps).

B. The *West Italic* subgroup

a) The *Sicilian* branch

b) The *Continental* branch, including, among several languages of small diffusion, *Latin*, in westernmost Latium, by the mouth of the Tiber and in the city of Rome. Except for Latin, which was to flourish as few other languages have, all the Italic languages died out before the end of antiquity and are known only through inscriptions. At the beginning of historical times, Latin was spoken over an area smaller than Zealand in Denmark, and it came to dominate half a continent. At the close of antiquity, it was continued in the *Romance* languages, under which form it can be

divided into six subbranches: (1) *Italo-Sardinian* (*Italian*, on the Italian peninsula, in Sicily, and, to some degree, in Corsica; *Sardinian*, in Sardinia and, to a certain extent, in Corsica); (2) *Hispano-Romance* (*Spanish*, in the greater part of Spain, in northwestern Africa, in all of America south of the United States except for Brazil, and in the Philippine Islands, *Portuguese*, in Portugal, Brazil, the Azores, and Madeira; *Galician*, in northwestern Spain); (3) *Gallo-Romance* (*Catalan*, principally in Catalonia and southeastern France; *Provençal*, in southern France; *French*, the official language of France and some former colonies, and one of the official languages (with Flemish) of Belgium, (with German, Italian, and Rhæto-Romance) of Switzerland, (with English) of Canada, and (with German) of Luxemburg); (4) *Rhæto-Romance* (in southeastern Switzerland and northeastern Italy); (5) *Dalmatian* (in Dalmatia, extinct since the nineteenth century); and (6) *Rumanian* (in Rumania, with the partial exception of southeasternmost Transylvania, with diffusion into the Balkans).

4. The *Hellenic* group

- A. The *Macedonian* subgroup, comprising only one extinct and poorly known language.
- B. The *Greek* subgroup, also comprising only one language.

5. The *Baltic* group

- A. The *North Baltic* subgroup (*Lithuanian* and *Lettish*). Lithuanian is the only living language that is immediately explicable from the Common Indo-European system of formulæ.
- B. The *South Baltic* subgroup, comprising several languages, of which only one is attested – *Old Prussian*, which died out in the seventeenth century, dislodged by German speakers who took over the name of the Baltic Prussians.

6. The *Slavic* Group

- A. The *South Slavic* subgroup
 - a) The *Bulgarian* branch (one language) and *Macedonian* (also one language, an older stage of which, in the form of *Old Church Slavic*, is the oldest attested Slavic Language).

- b) The *Jugoslav* branch (*Serbo-Croatian* to the east, *Slovene* to the west).
 - B. The *West Slavic* subgroup (*Sorbian* – also called *Lusatian* or *Wendish* – with two literary forms, the language of the Wends, in Lower and Upper Lusatia; *Polabian*, around the Elbe (Slavic *Laba*) and in Hanover, extinct since the eighteenth century; *Polish*; *Czech* and *Slovak*).
 - C. The *Russian*, or *East Slavic*, subgroup (*Great Russian*, *Ukrainian*, *White Russian*).
- 7. The *Albanian* group (comprising only one language).
- 8. The *Armenian* group (comprising only one language).
- 9. The *Iranian* group
 - A. The *West Iranian* subgroup (*Persian* in three stages: *Old Persian* in cuneiform inscriptions, *Pehlevi*, and *Modern Persian*, the official language of Iran; *Avestan*, indigenous in ancient times to the region north of Old Persian, attested in the writings of the religious leader Zarathustra, closely related to the younger *Chaldean Pehlevi*, and later extinct; *Kurdish*, in and around the mountain region on the border between Turkey and Iran; several other languages, including some in various regions of Iran).
 - B. The *East Iranian* subgroup (*Sogdian*, attested in texts from Chinese Turkestan, continued in the modern *Yagnobi*, which is spoken in the northwest of the Pamir, east of Lake Aral; *Sakian*, in texts found in Chinese Turkestan; *Afghan*, or *Pushtu*; *Ossetic*, in the Caucasus; various other languages including those of the Pamir and Baluchistan).
- 10. The *Indic* group. *Old Indic* comprises two dialects: *Vedic* (found in the Vedas) and *Sanskrit*. The *Middle Indic* languages are referred to under the general name of *Prākritis* and include *Pāli*, the sacred language of Southern Buddhism. Modern *Indic* comprises a very large number of languages (including *Romany*, the language of the Gypsies), spoken by some 400 million people. At the present time, Indic follows Germanic in number of speakers; next comes Romance, with a 370 million, and next Slavic, with about 270 million.
- 11. *Hittite*, an ancient language discovered only in recent times, recorded partly in hieroglyphic texts but mainly in cuneiform; indigenous to Asia Minor in the fifteenth century before our era.

12. *Tokharian*, discovered early in the present century, spoken in Chinese Turkestan in the seventh century of our era.

To the Indo-European family also belong some linguistic remnants (i.e. fragmentarily attested dead languages) of the Mediterranean region: *Messapic*, *Venetic*, *Thracian*, *Phrygian*, and *Lycian*.

II. The *Hamito-Semitic* family

1. The *Hamitic* group

A. The *Egyptian* subgroup, recorded in hieroglyphic inscriptions from the fourth millennium before our era. Younger stages are *demotic Egyptian* and *Coptic*, which died and was replaced by Arabic in the sixteenth century of our era.

B. The *Libyco-Berber* subgroup (extinct in Libya, replaced mainly by Arabic; modern languages in Algeria, Tunis, Morocco, and the Sahara).

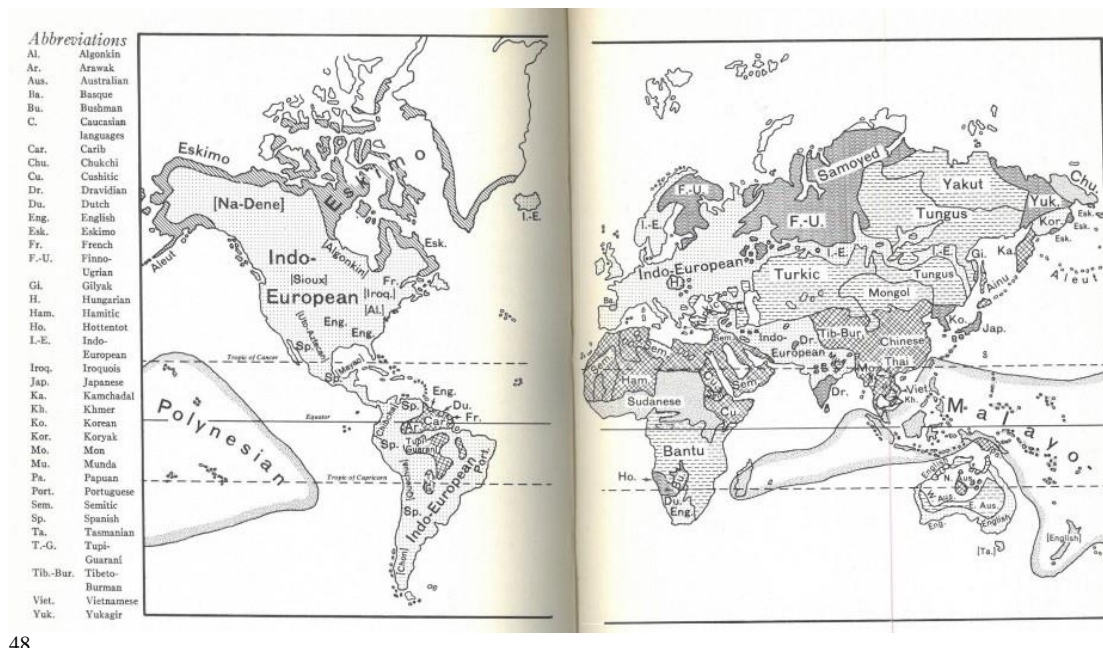
C. The *Cushitic* subgroup (eastern Nubia, Ethiopia, Somalia, East Africa).

2. The *Semitic* group

A. The *Akkadian* subgroup, comprising only one language – the ancient language of Iraq, recorded in cuneiform.

B. The *West Semitic* subgroup

a) The *Northwest Semitic* branch comprises two subbranches: *Cannanite* (*Old Canaanite*, *Moabite*, *Hebrew*, *Phoenician*) and *Aramaic*. With the exception of Hebrew and a few remaining dialects of East Aramaic, all these languages are extinct. Hebrew, to be sure, died out in Palestine as a spoken language in the third century before



our era and was replaced by Aramaic, but it continued to be a language of learning and religion, and it also found continuation in Yiddish, which can be viewed as a blend of Hebrew and High German. Moreover, Hebrew has been reintroduced as a living language, *Modern Hebrew*, in present-day Israel.

- b) The *Southwest semitic* branch comprises three subbranches: *Arabic* (which, as vehicle of Mohammedan culture, has been the most widely diffused of all the Hamito-Semitic languages; with English and Chinese, one of the most widespread modern cultural languages; spoken in modern times in Arabia, Iraq, Syria, Egypt, northwest Africa, and Malta), *South Arabic*, and *Ethiopic* (including seven modern languages, of which the best known is *Amharic*, the official language of Ethiopia).

- IV. The *Bantu* family (from *ba-ntu*, plural of *mu-ntu* 'human being') comprising a very large number of native languages in the whole southern half of Africa except for the southwestern corner.

⁴⁸ N.T.: Reprodução do mapa apresentado nas páginas 74 e 75 da edição em inglês.

V. The *Uralic* family

1. The *Samoyed* group, comprising five languages along the western part of the northern coast of the Soviet Union. Two of them acquired the status of literary languages in 1930.
2. The *Finno-Ugrian* group
 - A. The *Ugrian* subgroup
 - a) The *Ob-Ugrian* branch (*Ostyak*, on the river Ob'; *Vogul*, to the west of Ostyak, in the Ural Mountains; both languages acquired literary status in 1930).
 - b) The *Hungarian* branch (one language only, the official language of Hungary; also spoken in scattered parts of Slovakia and Jugoslavia).
 - B. The *Lapp* subgroup (one language only, but including six quite markedly divergent dialect groups, in Norway, Sweden, Finland, and the Soviet Union).
 - C. The *Finno-Permian* subgroup
 - a) The *Permian* branch (*Komi*, or *Zyrian*, in the Komi ASSR, in the northeastern corner of European Russia, with a rich literature from the fourteenth century; *Udmurt*, or *Votyak*, in the Udmurt ASSR, northeast of Kazan').
 - b) The *Finno-Mordvin* branch, comprising three subbranches: (1) *Mordvin*, a language spoken in scattered areas around the Volga between the cities of Gorky (formerly Nizhny Novgorod) and Kuybyshev (formerly Samara); (2) *Mari*, or *Cheremiss*, one language only, in the Mari ASSR (north of Kazan') and in scattered areas east of the Udmurt ASSR; (3) *Baltic Finnish*, comprising ten quite closely related languages, of which the best known are *Finnish*, *Karelian*, and *Estonian*.

VI. The *Altaic* family

1. The *Turkic* group, comprising *Turkish*, or *Osmanli*, in Turkey, in northern Syria, and in scattered areas of northern Greece and of Serbia; and a number of other languages up and down the Soviet Union – including the official language Azerbaidjan, of the Tatar ASSR, and of the Chuvash ASSR (*Chuvash*) and *Yakut*,

along the lower Lena – as well as in Sin-Kiang Province in Chinese Turkestan. The *Old Turkic* of the Orkhon inscriptions in Outer Mongolia was deciphered by Vilhelm Thomsen in 1893 and became the basis for the genetic comparison of the Turkic languages.

2. The *Mongol* group, comprising a number of languages mainly in Mongolia and the adjacent region of Siberia, together with parts of Kuku-nor Province. Literature dates from the thirteenth century.
3. The *Tungus* group (*Evenki*, in the greater part of Siberia east of the Yenisei, a literary language since 1930; *Manchu*, in Manchu-Kuo and the Soviet Dal'neVostochny Kray (between Nikolayevsk and Vladivostok), with literature beginning in the twelfth century).

VI. The *Sino-Austriac* family

1. The *Thai-Chinese* group, comprising *Chinese*, the language spoken by the greatest number of people in the world (600 million), and the *Thai* languages of Indo-China, of which the best known is *Siamese*, the official language of Thailand.
2. The *Tibeto-Burman* group.
3. The *Austriac* Group
 - A. The *Malayo-Polynesian* subgroup (Pacific islands, Malaya, Madagascar; Indonesia; the old cultural language of Java).
 - B. The *Austro-Asiatic* subgroup, comprising three branches: the *Munda* languages of India, especially in Bihar Province (the best known of these languages is *Santālī*), the *Mon-Khmer* languages of Indo-China, and *Vietnamese*.

Although we have used *family* as the most inclusive term for a large class of genetically related languages, there is, of course, nothing to prevent something we have called a family from being related to other languages. For one thing, it is possible to prove that two languages are genetically related, but it is never possible to prove that two languages are *not* genetically related. We use the term *family*, therefore, for the largest class of related languages that we can set up at a given stage of our research. In this respect, family is a relative term, and if two families are set up which are later found to be genetically related, we should, if you will, call them one family, each of them a group, and so on. But this would usually be impractical. Even though, for example, we now know that the Indo-European languages are genetically related to other languages, we still

continue to speak of the Indo-European family, and we may very well do so, provided only that we admit the possibility of there being *families of families*.

In each case, it will be merely a practical question whether we change our terms or not. For a long time, what we have called the Sino-Austic family was seen as five different families having no demonstrated genetic relationship: (1) Sino-Tibetan, comprising Thai-Chinese and Tibeto-Burman, (2) Malayo-Polynesian, (3) Munda, (4) Mon-Khmer, (5) Vietnamese. In 1889 the German linguist Ernst Kuhn demonstrated that the last three are related, and in 1908-10 the Danish linguist Kurt Wulff was able to show their relationship to Malayo-Polynesian. Much later, the genetic relationship between Austic and Sino-Tibetan was finally demonstrated in a posthumous work of Wulff's. We have drawn the consequences from this and combined all these groups into one family, because the genetic relationships that Wulff demonstrated contradict the older classification: the relationship between Thai-Chinese and Austic has been shown to be closer than that between Thai-Chinese and Tibeto-Burman, so there is no longer any basis for maintaining a separate Sino-Tibetan family.

A genetic relationship between Indo-European and Hamito-Semitic was demonstrated in detail by the Danish linguist Hermann Møller, using the method of element-functions, in studies dating from 1906-17. Later, by the same method, the French Africanist Mlle Lilius Homburger thought it possible to demonstrate genetic relationship between Hamito-Semitic and Bantu and even conjectured adding to the family both some northeastern (Nilotic) and some western Sudanese languages, of which the northeastern seem to be nearer than Coptic to Old Egyptian and demotic Egyptian. And there can be no doubt of the genetic relationship between Indo-European and Uralic after Holger Pedersen's success in demonstrating a number of element-correspondences.

By combining these far-reaching results we may perhaps reduce the six families listed above to three: (1) *Nostratic* (a name proposed by Holger Pedersen for the languages related to our own, from Lat. *nostrās* 'our fellow countryman'), comprising, so far as we can see at present, Indo-European, Hamito-Semitic, Uralic, and possibly Bantu (together with the Sudanese languages mentioned above); (2) *Sino-Austic*, (3) *Altaic*.

The language families discussed here are the only language classes of any considerable size to have been established as families by the method of element-functions. The same method has been applied to the families of families that we have talked about in the last paragraph, but because the great syntheses involved are so recent, we must

prudently admit that proof has not yet been given, even in those instances where there is no doubt that it will be.

Taken together, the languages we have listed above are spoken by about eight-ninths of the world's present population. But linguistics must, of course, be concerned with dead languages as well as the living.

The remaining languages of the world constitute an enormous number of genetically isolated speech communities or families of quite restricted size. Examples are Japanese, Korean, Ainu in northern Japan,⁴⁹ the Palæo-Asiatic languages of northeastern Siberia (a purely geographical classification), South Caucasian (a true language family), North Caucasian (rather a purely geographical classification of a very large number of scattered languages), Davidian (in southern India, a true family), the Australian languages (merely a geographical classification), the non-Bantu languages of the Sudan, the Khoisan family in southwest Africa (Bushman and Hottentot), Basque (an isolated language on the French-Spanish border), Eskimo (together with Aleut), and, finally, the extraordinarily numerous American Indian languages, which are still classified into no fewer than some hundred different families. To these may further be added a number of isolated linguistic remains, principally in the Near East and the Mediterranean region (for example, Sumerian in Iraq, and Etruscan in Italy).

Parent Languages

We have given an operational definition of genetic relationship above (p. 30) and shown that it consists in what we have called element-functions, constant correspondences between the systems of expression elements in the related languages. Genetic relationship is demonstrated by the demonstration of such element-functions, and in no other way. A superficial similarity between languages or between signs in different languages plays no role in genetic relationship or in its demonstration. Such "similarity" cannot be used as a scientific criterion, if only because a feeling that things are similar is a purely subjective matter (cf. above, p. 43). In many instances, two signs can be shown to be genetically related even though most people would perhaps find no resemblance between them. Such forms as the nominative singular *hayr* in Armenian and *patér* in Greek, or the genitive singular *hawr* in Armenian and *patrós* in Greek, would probably

⁴⁹According to recent investigations, Ainu would appear to be an Indo-European language.

not generally be taken as having the least external resemblance, yet they can be shown, on the basis of element-functions, to be photographically identical. On the other hand, two signs may bear a strong resemblance without having any genetic relationship whatsoever. Primary examples of this are most loan words, and if we should attempt to base our constructions on external similarity, it would be impossible to distinguish between borrowed and genetically related words. But of course there are also instances of two signs in different languages resembling each other by pure accident: 'big' is *dīdis* in Lithuanian and *didi* in Georgian, but the two languages are genetically unrelated and the two words have nothing to do with each other; to *call* is *kaleîn* in Greek, but these cannot be the same word since Gr. *k* in this position corresponds genetically to Gmc. *h* (cf. *kúōn* and *hound*, *he-katón* and *hund-red*, etc.).

Now if we ask for the cause of this element-function that we have called genetic relationship between two languages, we see that it can be accounted for only on the assumption that the two languages have a common origin. This is the only conceivable explanation, so we are logically compelled to accept this consequence of genetic relationship. Genetic relationship consists in a connexion between the expression elements and signs of the two languages, and the reason for this connexion must necessarily be that the expression elements and signs of the two languages are later developments of older expression elements and signs that were common to the two languages, or, in other words, that the two languages developed from the splitting of a single *basic language* (commonly called *parent language*). Thus (according to the classification given in the preceding chapter), behind Danish and Swedish there must lie an East Scandinavian parent language, behind this and West Scandinavian a Scandinavian parent language, behind this and Gothic an East Germanic parent language, behind this and West Germanic a Germanic parent language, behind this and the other Indo-European language groups an Indo-European parent language, and behind this and the language families that are genetically related to Indo-European a Nostratic parent language. Now we have a solid ground of reality under our feet, and we can supplement our earlier operational definition of genetic relationship with a "real" *definition: Genetic relationship between two languages means that they have developed from a single basic, or parent, language.*

All that we can know with any certainty about the parent language is what we can read from the formulae we have established on the basis of genetic comparison of languages. In no instance has the parent language been transmitted to us, we *reconstruct*

it from the element-functions. The parent language that we presuppose for the Romance languages may well be called Latin, because we can conclude from our knowledge of history that it *must* have been Latin which, by splitting, developed into the Romance languages. But, as we have already noted (p. 22), the parent language that we are led to reconstruct for Romance is not the same Latin as the Latin that we know from the transmitted literature.

The expression elements and parts of words (almost never whole words; see p. 27) that we set up in our formulæ must, then, be ascribed to the parent language. They are fragments of a language that we know in no other way. And here it is very important to realize clearly how much and how little real content we have a right to read into our formulæ – in other words, what we know with certainty and what we do not know with certainty about the parent language:

A. We know with certainty

1. that such a language existed;
2. that the language contained in its structure the expression elements represented by our element-formulæ (provided, of course, these formulæ have been correctly set up – not a completely superfluous proviso, as we shall shortly see);
3. that these expression elements in the structure were defined by the possibilities of combination that are implied in our formulæ for signs.

B. We do not know with certainty

1. where, when, or by whom the language was used, or whether it was ever used (one possibility to be reckoned with is that it was never used, but was transformed from the very start in the directions implied by the several languages under comparison);
2. whether it contained in its structure more expression elements than those designated by our element-formulæ;
3. whether the expression elements we know to have been in the language were defined in the structure by more possibilities of combination than those implied in our formulæ for signs;
4. how the linguistic usage (if there was one; see point 1) of the language was organized and, in particular, (a) what whole words or sentences were, or could be, used, (b) how the expression elements were represented in speech, in writing, or in other ways, and (c) what meaning was contained in the signs.

In these last four respects, where we have no certainty, we may, if we wish, construct *hypotheses* based on probabilities. It cannot be denied that such hypotheses are extraordinarily tempting and that, to a certain extent, they arise quite naturally. Here we have a happy hunting ground for romantically disposed and imaginative spirits. By combining probability and fantasy, we can conjure up a whole picture of a prehistoric language like the Indo-European parent language, with words and sentences and a definite pronunciation and definite word-meanings, constructed to “resemble” as closely as possible words, sentences, pronunciations, and meanings in the historically known languages. And from this, another picture can be conjured up of a whole prehistoric society and prehistoric culture, all based on conjectures about what words existed and what things they designated. And, still hypothetically, this prehistoric society can be assigned some definite place or other on the basis of conjectures, say, about the plants and animals that might have been designated by the words in the language. Naturally, science has not refrained from constructing and discussing such hypotheses, quite justifiably – so long as it is remembered that they can never be anything but hypotheses. Classical linguistics, however, was well on the way to making such hypothesis-spinning the essential part of genetic comparison and thereby contributed to bringing it into unjust discredit as an insecure and hypothetical science. It is just the opposite. It is a science with secure results and an exact method. But it can maintain itself as such only by keeping clear of the hypothetical domain, cleaving to what is known with certainty, and respecting the limitations of its method. In no way does the method permit us to infer culture, people, or “race” from language. The fact that our language is Indo-European can, of course – if we care to use an inappropriate wording – be expressed by saying that we are Indo-Europeans. But no other meaning can be correctly attached to this sentence than that our language is connected with other Indo-European languages by element-functions. This means that the system of elements in our language and bits of its words go back to an earlier, purely linguistic, Indo-European unity, but it in no way means that we ourselves go back to some unity connected with it. Precisely because languages can spread and move and be transferred to quite different populations from their original speakers (as Latin, starting from the little area by the mouth of the Tiber, was disseminated over an enormous number of populations differing greatly in all respects, or as, in more recent times, English has spread to many quarters of the world), it is impossible to maintain any such connexion between languages and peoples. There is not even any certainty that the

possibly quite small population that spoke a language thousands of years ago (like, for example, the small band of Latin-speaking men by the mouth of the Tiber) actually has any direct descendants now or, if it does, that those descendants speak a language genetically continuing the language of their ancestors. This element of uncertainty obviously becomes greater the more language families are involved and, consequently, the farther back we may go into prehistory in our hypothetical dating of the parent language. Linguistic arguments, therefore, cannot be used to support racial theories.

As we have seen, the element-functions are set up without regard for the way the elements are represented in the languages under consideration (whether, for example, by letters or by sounds), and the element-formulæ must therefore likewise have an abstract character. It is certainly true that the elements we compare in the different Indo-European languages, for example, are represented by letters: after all, in the case of dead languages like Ancient Greek and Latin, Old Indic, and Gothic, we have nothing but written characters to deal with and can only construct more or less probable hypotheses about their pronunciation. But if we do not argue from the letters of these languages that the Indo-European parent language also had letters it is not merely because we happen to know that the Greek and Latin alphabets (and the Indic) are transformations of the Semitic alphabet and consequently have a non-Indo-European origin. Rather, it is above all because our method simply will not permit us to say anything whatsoever about how the elements of the parent language were represented. The same is of course true (one might be tempted to say, to an even higher degree) when it comes to the representation of elements by certain sounds in pronunciation. In this matter, only hypotheses can – and no hypotheses need – be offered. Such hypotheses may have a greater or lesser degree of probability, but it is always a question of probability only, never of certainty. The hypothesis that IE **m* was pronounced approximately like our *m* has a very high degree of probability; the probability that IE **ʌ* was pronounced approximately like the last vowel in our word *sofa* (one of the hypotheses that have been suggested) is very low. It is therefore unfortunate that classical linguistics called the element-functions *sound laws* and, for example, called the consonant-functions discovered by Rask for Germanic (p. 19) and High German (p. 24) the Germanic and High German *sound shifts* (Rask himself called them *letter changes*, a considerably more honest and sober term).

In the same way, specifications of meaning for the words (or, better, the parts of words) of the parent language remain hypothetical and, like specifications of pronunciation, are superfluous and irrelevant from the point of view of the theory. Here

too, the degree of probability can be greater or less. Languages seldom show such agreement with respect to meaning as they may show with respect to the pronunciation of an element (where the pronunciation in the several languages is known or can be determined with a high degree of probability); very rarely do we find such unanimity over the meaning of a word as we have over the pronunciation of the element **m*. It might be thought that there is a high degree of probability in the assumption that IE **patḗr* meant ‘father’. But in the first place, **patḗr* is composed of a root **p_A-* and a derivational suffix **-t-r*, which contains yet a third component, the vowel **ē* that alternates with **ō* (see p. 29). The root **p_A-* is in regular alternation with another form, **pō-*, which appears in many Indo-European languages (e.g. in OI *pá-ti* ‘he protects’) with meanings like ‘protect’, etc. The suffix **-t-r* is known from many other words. In the several Indo-European languages it designates the person who carries out an action (e.g. Gr. *rhḗ-tōr* ‘speaker’, Lat. *ōrā-tor*). In other words, **pa-tḗr* is a regular formation that we might expect most probably to have meant something like ‘the protector’. It need not have existed in the parent language; it is so regular and obvious that it could have been formed independently in the several Indo-European languages at an early date. And whether or not it belonged to the parent language, there is nothing to indicate that its original meaning was ‘father’.

One reason why the content of meaning must necessarily remain hypothetical is that, as a rule, we are not able to reconstruct whole words but only *minimal signs*, i.e. the smallest word-components (such as roots, suffixes, and flexional endings that can be shown to be meaningful in various languages. The moment we are faced with a formation of regular type in Indo-European, we have to admit the possibility of its having arisen at any time, in the parent language or later, just as the English ending *-er*, denoting an agent, can be freely added to any verb, even if it never has been before. Corresponding to the Danish verb *leje* ‘to hire’ (Goth. *leihwan*) we find Gr. *leípein* ‘to leave’. The Greek present *leípō* is to be understood as IE **leik₃-ō*. In Latin, the present is *linquō*, equaling IE **li-n-k₃-ō* (in which the **e* of the root is alternating with zero – see p.29 – and an *n*-component has been infixéd into the root). In Old Indic, the present of the same verb is *riṇácmi*, equaling IE **li-né-k₃-mi* (with a different ending, *-mi*, for the first person singular and with a different alternation grade in the infixéd component). Now all these formations represent *possible* types of present formations for such a verb in Indo-European, but we cannot say which of them was used in the parent language – possibly none, possibly one or more.

Thus we reconstruct, not words, but parts of words, and without knowing their pronunciation or their meaning. This might seem to be pitifully little – but in fact it is a very great deal. By using the little information that we have, we can prove genetic relationships within very broad, indeed unlimited, linguistic domains; and we can say a lot about the structure of the parent language and thereby explain a great number of details in the various languages belonging to the family. The method even permits us to *predict*. It permits us, for example, to infer with certainty that if the word *mother* existed in Gothic, and if it was not subjected to action of elements, analogical transformation, abbreviation, or tabu, it must have been *modar* – of that we can be quite certain. In other words, reconstruction can be reversed, and we can thus construct for ourselves the form that a word must have had in the individual languages within the language family. Such reconstruction is, in fact, one of the prerequisites for distinguishing between borrowings and genetically related words, as well as between element-functions and counter-examples.

Genetic linguistics may therefore be called absolutely exact, and is perhaps the most exact domain of the humanities.

As we have already noted, only a minority of the words in a language can, in whole or in part, be shown to have genetic connexions with words in related languages – whether because of new formations, borrowings, or transformations. Only within this small sector of the vocabulary can genetic linguistics use the method of element-functions. It is therefore quite impossible for genetic linguistics to account for all the words in a language by means of its exact method. Even if we must always try to explain as many words as possible in this way, always respecting the element-functions, genetic linguistics can only to a very limited extent give an unambiguous and certain answer to questions about *etymologies*, i.e. the individual histories of particular words. In most cases this is an extremely complicated problem, without any unambiguous solution, for which the method of genetic linguistics is inadequate. An etymological dictionary sets itself the tempting, but in fact unscientific, task of accounting for *all* the words in a language. In doing so, it strains the method far beyond its powers and brings unmerited discredit on it. In order to estimate the probability of the etymological hypotheses such a dictionary offers for the origin of a word, one must first know the element-functions within the language family concerned, so as to be able to see whether the hypotheses agree with them. A layman who has not studied the element-functions cannot, in fact, use an etymological dictionary with reasonable and fair judgment.

Typological Relationship of Languages

We observed earlier that it should be possible to set up functions between languages other than the element-functions. In addition to being ordered into families, languages may be ordered into linguistic *types*, according to their structure. Such classification depends, of course, on what structural features are used as criteria. Just as men can be divided into tall and short, white, black, yellow, and red, dark-haired and blond, etc., so the universe of languages can be divided in a number of different ways, depending on one's point of view. One possible choice is between classification according to linguistic structure and classification according to linguistic usage, with various bases for further division under each case.

The only linguistic typology to achieve a place in classical linguistics was a classification according to linguistic usage. The central point of interest was the structure of signs, especially of words. *Words* are permutable signs, signs that can exchange places within a linguistic chain: *softly answered* consists of two words, because one can also say *answered softly*, *soft-ly* and *answer-ed* each consist of two signs, but these signs cannot be put in another order. Permutable signs attracted an extraordinary amount of attention from classical linguistics, beginning with antiquity, since it was thought, in connexion with Aristotelian conceptual logic, that each such sign stood for one concept. This also explains the interest that has been felt in dividing the universe of languages according to the word structure chosen by the linguistic usage.

As a rule, four linguistic types have been distinguished:

1. An *isolating* type, in which the words are invariable, not subject to inflexion or derivation, and in which grammatical relationships can be expressed only by word order (as in English *dog bites man* and *man bites dog*, where different word order alone expresses grammatical relationships expressed in other languages by different cases) or by the addition of independent words (for example, the plural may be expressed by adding the word 'many' to the form that also serves as singular, or the present tense by adding the word 'now'). The example always quoted for this type is Classical Chinese (modern spoken Chinese is a different matter).

2. An *agglutinating* type, in which all grammatical relationships (derivations and inflexions) are expressed by suffixes, signs attached *after* the lexical component or root of the word, with each relationship having its own special suffix. In Turkish, where

the suffix *-lar-* expresses the plural, *-a* the dative, *-da* the locative, and *-dan* the ablative, the word *kuş* ‘bird’ is declined as follows:

	singular	plural
nominative	<i>kuş</i>	<i>kuş-lar</i>
dative	<i>kuş-a</i>	<i>kuş-lar-a</i>
locative	<i>kuş-da</i>	<i>kuş-lar-da</i>
ablative	<i>kuş-dan</i>	<i>kuş-lar-dan</i>

The Altaic and Uralic languages are taken as principal examples of this type.

3. An *inflecting* type, in which the boundary between root and suffix is not clear, in which an individual suffix often expresses several different grammatical relationships at once (like Lat. *-ārum* in *bonārum*; see p. 32), and in which the root itself may undergo change during inflexion, with vowel alternations or inserted components. The older Indo-European languages have been taken as models of this type.

4. A *polysynthetic* type, in which all grammatical relationships in the sentence can be expressed by affixes or by alterations of a single root. A sentence often is equivalent to a single word, none of its components being permutable. Greenlandic has been cited as an example: a sentence like *kavfiliorniarumagaluarpunga* ‘I should like to make a coffee’ contains only one word.

Many objections have been made to this classification. One of its unsatisfactory features is that the types are almost never found pure, and most languages represent mixtures. The classification was originally set up by Friedrich von Schlegel and Wilhelm von Humboldt. In our times, the American linguist Edward Sapir developed it further, in a very ingenious but extremely complicated way, into a system intended to provide place for all conceivable word structures. In this supplemented and clarified form, the typology of word structure merits consideration. We shall not use Sapir’s elaborate schema here, but shall only consider briefly what can be learned by setting up such linguistic types.

Although the original attempt was to identify the linguistic types with certain families – the isolating type with Thai- Chinese, the agglutinating with Altaic and Uralic, the inflecting with Indo-European, and the polysynthetic with certain American languages, for example – it soon became evident that the attempt could not succeed. However linguistic types are set up, they will clearly never coincide with linguistic

families. The very fact of linguistic change, which underlies the emergence of language families, implies that a language may change type in the course of time. Even if it can be maintained that Indo-European was once inflexional, many Indo-European languages have now passed over to quite other types, without, of course, ceasing to be Indo-European. An Armenian declensional schema like

	singular	plural
nominative	<i>hay</i> ‘an Armenian’	<i>hay-er</i>
dative-genitive	<i>hay-u</i>	<i>hay-er-u</i>
instrumental	<i>hay-ov</i>	<i>hay-er-ov</i>
ablative	<i>hay-ē</i>	<i>hay-er-ē</i>

obviously illustrates pure agglutination. It has been supposed that this is due to Turkish influence; however that may be, Armenian, even in its oldest form, no longer belongs to the inflecting type.

Another Indo-European language, Modern English, has, if anything, become isolating in type and in this respect is closer to Chinese than to its genetic origin. The same has been asserted of Modern French by some linguists, but in the opinion of others that language is instead on the way towards polysynthesis: in the sentence *je ne le lui ai pas donné* ‘I didn’t give it to him’, only one permutation is possible (*ne le lui ai-je pas donné?* ‘Didn’t I give it to him?’), so we may here be faced with the kind of sentence-words we found in Greenlandic.

Thus, genetic and typological relationship are two quite different things that have nothing to do with each other. Within one and the same family, we can find languages of fundamentally different types, and within one and the same type, languages of fundamentally different families. The two divisions of the universe of languages have nothing in common. They are established on completely different bases and from completely different points of view. In the same way, the zoologist can classify animals systematically as mammals, fish, etc., but also ecologically, setting up, for example, a category of marine animals that would include whales and seals, octopuses, oysters, and fish – animals belonging to quite distinct categories in the systematic classification.

Classical linguistics has been far less interested in typological than in genetic linguistic relationship. But no long reflexion is needed to see that this is a rich and extremely important field, although still almost entirely uncultivated. It would clearly be

possible to set up a linguistic typology from many other points of view than the one that was arbitrarily chosen by classical linguistics and that turns out to be the most superficial of all – that of word structure. It is also easy to see that, whatever points of view are chosen, the situation will remain the same: the resulting classification will cut across the genetic classification and be quite unrelated to it.

When we try to set up a linguistic typology from more penetrating and immanent points of view than have been chosen till now, we soon come to see that the entities which must be compared in the different languages are *categories*. These will be (1) categories in the linguistic structure (like vowels and consonants, accents and modulations, and their various subcategories on the expression side, and grammatical categories like case, gender, and number on the content side) and (2) categories in the linguistic usage (like the expression categories of sounds and the content categories of meanings). From all these points of view, the languages that show the same categories must be assigned to the same linguistic type. Comparison of elements, on the other hand, will never be of interest for linguistic typology, where the fact that two languages have an element in common will always mean that they have a category in common – the category represented by that element. This will come out more clearly from what follows, but we are already in a position to state that this is why genetic relationship and typological relationship remain two essentially different things. Having operationally defined genetic relationship as a function between languages that consists in element-functions, we may give a corresponding *operational definition of typological relationship as a function between languages consisting in the fact that categories in each language have function to categories in each of the others*. Just as genetic relationship is recognized on the basis of element-function, typological relationship, when it penetrates below the most superficial and accidental level (that of word structure), is recognized on the basis of *category-function*.

Such a typology will also naturally account for the structure of the units (syllables, for example) that enter into a language, since the structure of these units depends simply on the categories found in the language. We have defined a category as the sum of the entities that can be inserted at definite places in the chain (p. 33), and we have shown that vowels and consonants are examples of categories (p. 35). A unit like the syllable, then, will depend on the existing categories of elements, and so for all other units.

An exhaustive linguistic typology is, in fact, the biggest and most important task facing linguistics. Unlike linguistic genetics it has no regional limitations. Its ultimate aim

must be to show which linguistic structures are possible, in general, and why it is just those structures, and not others, that are possible. And here it will come closer than any other kind of linguistics to what might be called the problem of the essence of language. Finally, it will prove to be superordinated to genetic linguistics, since only through linguistic typology can we hope to understand what laws govern linguistic change and what possibilities of change a given linguistic type implies. Only through typology does linguistics rise to quite general points of view and become a science. At present, to be sure, very little has been realized of this high hope. The task has been set, but is still far from being accomplished. Classical linguistics has been seriously interested only in usage and genetics, while structure and typology have been left almost completely uncultivated. The central problems have not been solved – have not even been posed – by classical linguistics. This is deplorable, if you will, and the reader may well deplore the fact that the following exposition of linguistic typology will necessarily amount to nothing more than sketchy suggestions – an unrealized program, a questionnaire with unfilled blanks. But for the scholar there is nothing more beautiful than the vision of a science still to be created.

Types of Linguistic Structure

Since a category consists of the entities that can occupy a given place in the chain, we must start from the chain as the immediate datum in determining the structure of any particular language (*état de langue*); it is through an analysis of the chain that we arrive at the categories. A whole linguistic chain that is so taken as an object of analysis is called a *text*.

Naturally, the analysis must not be a haphazard chopping up of the text but must be performed with an eye to the *relations* between its constituent parts (see p. 32). These relations may be of several kinds.

First, we have to distinguish between *implicational* and *nonimplicational relations*. We shall say that two entities have an *implicational relation* whenever the presence of one of them in the chain is a necessary presupposition for the presence of the other, or, to put it the other way, whenever one of the entities necessarily implies the other. And we shall speak of a *nonimplicational relation* when there is no such presupposition between the related elements – when we can have either entity present without the other.

Nonimplicational relation is very common between elements. In *bland* there is a relation between *b* and *l*: they can appear together in the initial field of the syllable. But the relation is nonimplicational; we find *l* without *b* (*land*), and we find *b* without *l* (*band*). As a rule, we can also find such a nonimplicational relation so long as we are considering only some particular consonant in its relation with some particular vowel. There is a relation between *b* and *e* in the syllable *be*, but the relation is nonimplicational: we can have a syllable with *b* and without *e* (*by*) and a syllable with *e* and without *b* (*me*). It turns out, however – after we have established the categories on the basis of the textual analysis – that there will be an implicational relation between the category of vowels and the category of consonants within the syllable: we cannot have a representative from the category of consonants without having a representative from the category of vowels in the same syllable (cf. p. 35). In the same way, there is nonimplicational relation between a particular subordinate clause and a particular main clause, since the subordinate clause can be replaced by another subordinate clause and the main clause by another main clause; but there is an implicational relation between the category of subordinate clauses and the category of main clauses. It is possible, however, to find implicational relations

between particular elements, and, as we shall see shortly, there are many examples to be found of these, both in everyday life and in the world of language.

Implicational relations may be further classified as *reciprocal* or *unilateral*. There is a reciprocal implicational relation between two entities when each presupposes the other; when one presupposes the other, but not vice versa, the relation is unilateral. Examples of *reciprocal implicational relations* are the relations between volume one and volume two of a book, between marriage partners, or between two betrothed – they could not be called what they are unless both existed. Let us add at once: if an author writes a book and calls it volume one, he has, by calling it that, presupposed the existence of volume two, even if that volume never comes out. In the same way, if a man is called a husband, the existence of his wife is presupposed, even though we might never happen to meet her. In other words, we may be given only one entity together with an implicational relation, and on that basis we can *supply* the relation with another entity not with any particular concrete entity that we could describe in detail, but with an entity about which we know only that it is something presupposed by the first. In such a case, where a presupposed entity is not manifested to us but can be supplied we say that it is *latent*.

If a language is so constructed as to have no syllables containing a vowel alone, but only syllables containing both vowel and consonant, we say that there is a reciprocal implicational relation between the category of consonants and category of vowels in the language. But if the syllables of a language may consist of vowel alone or of vowel and consonant, we conclude that the language displays a *unilateral implicational relation* between the category of consonants and the category of vowels: consonant *implies* vowel, but not vice versa; vowel is *implied* by consonant, but not vice versa. Unilateral implicational relation is the usual relation between the category of subordinate clauses and the category of main clauses: subordinate clause implies main clause, but not vice versa; main clause is implied by subordinate clause, but not vice versa. And if we encounter a subordinate clause all by itself without any trace of its main clause, we conclude that the main clause is latent and supply it as needed – not a particular concrete main clause, but one about which we know only that it is a main clause implied by the given subordinate clause. An exclamation like *If I had money!* requires such a supplementation. Examples of unilateral implicational relations are, of course, easy to find outside language as well. In chess, the king is implied by all the other pieces – if the king is mated, all his men are captured and the game is over – but not vice versa.

To be certain of securing an exhaustive description of a language, such that all relations are registered, we cannot straight away break up the text into the smallest elements having demonstrable relation, but must subject the text to a series of successive divisions such that at each step there will be as few parts as possible. In this way, any linguistic text can first be divided into two parts, a *content plane* and an *expression plane*, with reciprocal implicational relation: if it really is a language that we are dealing with, it must have both a content and an expression. Each of these is then further divided into parts and into parts of parts – chapters, paragraphs, sentences, clauses, etc. – down to the smallest elements.

At each step of this division we can set up categories defined by one or another of the relations we have been discussing. One category – for example, that of subordinate clauses or that of consonants – will be defined as implying; another – for example, the category of main clauses or that of vowels – will be defined as implied; and so on. Now an interesting fact is that the members of each such category have the peculiar property that the substitution of one member for another can entail a difference in the opposite plane of language. The expression elements for example, that are members of the consonant and vowel categories will each be defined as differing from the other members of the same category by the fact that substitution of any of these others for it can entail a difference in the linguistic content (in the meaning of a sign). The fact that *p*, *s*, *r*, and *t* are four different expression elements in English and that *a*, *e*, *i*, and *o* are four different expression elements in English can be seen at once from the fact that the exchange of one for another at the same place in the chain entails a difference in the content: *pat*, *sat*, *rat*, *tat*, *pat*, *pet*, *pit*, *pot* (cf. p. 33). It is with the help of such a *commutation test* that we are able to determine the number of members in a linguistic category. The same, obviously, holds true of the content elements of language: ‘nominative’ and ‘genitive’, ‘present’ and ‘preterite’, as well as all lexical components, are linguistically different by virtue of the fact that exchanging one for another can entail a difference in the expression. And it is also obvious that this test is applicable not only to what we have here called expression elements and content elements, but to all members whatsoever of all categories whatsoever. For example, two sentences are different only if their contents are different and their expressions are different. If we merely change the expression of a sentence, in such a way as to involve no difference in its content, we have produced only a difference in pronunciation, not a linguistic difference; and if we merely change the content of a

sentence, in such a way as to involve no difference in its expression, we have produced only a difference in thought, not a linguistic difference.

We have arrived at a fundamental property of language, one by which is distinguished from other structures. Taking the word *commutation* in the special sense in which we have used it, we can give a concise definition of *language* as *a structure in which the members of each category have mutual commutation*.

Clearly, this definition is a very broad one; and when we apply it we must be prepared to discover that it is satisfied by many structures that we are perhaps not accustomed to calling languages, at least not ordinary everyday languages. So, for example, it will turn out that everything we call uniforms, in the broadest sense, including folk costumes and the like, will constitute systems that fit this definition. Every uniform has a corresponding content: this uniform “means” a French officer of a certain rank; this vestment “means” a Roman Catholic prelate of a certain rank; and when, as very often happens, there are different folk costumes for young girls and for married women, each of these “uniforms” likewise has its special “meaning.” And the commutation test can be applied here as well: different content corresponds to different expression (different costume), and vice versa. And there will be a number of other areas where we shall find structures proving to be, on this decisive point of commutation, identical with the structures that we are used to calling languages. In so far as they attempt to communicate a content, the various art genres, like descriptive painting, program music, and so on, must be languages in the sense of our definition. A game, if interpreted, will also be a language in this sense – as, for example, a game of chess if it is interpreted as two battle arrays, each consisting of higher and lower ranking officers and privates and, among the higher rank officers, of king and queen, and so on. In such a case we have both content and expression and can apply the commutation test.

At the same time, we can appropriately narrow our definition somewhat by adding a qualification. If we look at our everyday languages, we shall always find in their structure the peculiar feature that a single expression element does not, as a rule, correspond to (i.e. does not have relation with) a single content element, and vice versa. As we have seen earlier (p. 32), it *may* happen that one expression element will correspond to one content element (e.g. Eng. *-s* as expression of ‘genitive’), but this is a special case and is never true of the whole structure of a language. As a rule, a sign consists of two or more expression elements and two or more content elements connected with them. Now as far as the linguistic expression is concerned, we are so fortunate as to

know – or to imagine that we know – how many expression elements enter into a given sign. As far as the linguistic content is concerned, at the present stage of research we still lack sufficient information. But we can easily perform a provisional analysis that will be adequate to our present purpose. The expression of written English words *ram* and *ewe* can be analyzed into three elements each: the letters *r*, *a*, and *m*, and *e*, *w*, and *e*. As for the content, each word can be analyzed into two elements: ‘he-sheep’ and ‘she-sheep’ respectively. We could certainly carry the analysis further by similarly analyzing the word *sheep* (and perhaps also the words *he* and *she*) into components, but that would take us too far for our present purpose. As can already be seen, what we are doing in this content analysis of *ram* and *ewe* is replacing the simple sign-content with a compound content whose components enter into other sign-contents in the language. (Thus, the components ‘he’ and ‘she’ enter into the content of the signs *he* and *she*, *boy* and *girl*, *stallion* and *mare*. To what extent the component ‘sheep’ enters into any compound sign-contents we shall not here inquire, since the answer would require the analysis of ‘sheep’ that we have deliberately not attempted; at any rate, the component ‘sheep’ obviously enters into the sign-content of the word *sheep*.) This replacement of a simple sign-content by a compound does not involve any change in the expression: it is all the same whether we say that *ram* means ‘ram’ or that it means ‘he-sheep’. What we have done here with the content corresponds completely to what we are better used to doing with the expression. When we resolve *ram* into *r*, *a*, and *m*, we are replacing the simple (i.e. still unanalyzed) sign-expression with a compound consisting of components that enter into other sign-expressions of the language. Again, this replacement of a simple sign-expression by a compound does not involve any change in the content: it is all the same whether we say that the concept ‘ram’ is expressed in English by *ram* or by *r*, *a*, and *m*. In both instances, in the content as well as in the expression, we *define* the unanalyzed entity by assembling other entities taken from the same language. This is always the way a definition is made, and the definitions of meaning that are found in dictionaries are to be understood as attempts at the kind of content analysis we have outlined. To avoid misunderstanding, we must also observe that a definition obviously does not consist of a helter-skelter enumeration of the sign components, but posits a special kind of relation between them. The expression definition of *ram* must include the information that *r*, *a*, and *m* appear in just that order, and, correspondingly, the content definition posits a certain logical relation between the components into which the sign-content is analyzed.

can be translated. Every game of chess can be translated into – formulated in – an everyday language, but not vice versa. In general, an everyday language differs from all other kinds of languages (e.g. the mathematician’s symbolic language or the chemist’s language of formulæ) by not being made especially for particular purposes but being of use for all purposes. In an everyday language we can – with the help of circumlocutions, if necessary, or carefully prepared explanations – formulate anything whatsoever. Any piece of program music, even, will be translatable into a piece of everyday language – but not vice versa. In everyday language as Søren Kierkegaard has said, one can “work over the inexpressible until it is expressed.” This is the advantage of everyday language, and its mystery. And this is why the Polish logician Tarski (who reached the same conclusion independently of the present author) rightly says that everyday languages are characterized in contrast to other languages by their “universalism.”

To establish a general theory of linguistic structure, we must start with some definition of language of the kind we have been considering, because the establishment of such a theory is not simply, or even principally, a matter of experience, but rather of calculation. Experience is no adequate basis for a theory of linguistic structure: it would be impossible to go through all existing linguistic texts, and, moreover, it would be futile, since the theory must be valid not only for the texts that have been written or spoken up to now, but also for those that will be written or spoken in the future – that is to say, for all theoretically possible texts and all theoretically possible languages. On the basis of certain experiences, which must necessarily be limited even if they should be as varied as possible, we make a calculation of all conceivable possibilities within certain frames. These frames we set arbitrarily, by taking – as we have just now done – certain features that are present in all the objects commonly called languages and generalizing so as to say: all objects displaying these features I shall call languages. It may turn out that some of the objects falling within the definition are not commonly called languages, but that makes no difference provided only that no objects are excluded that *are* commonly called languages. After we have thus defined – arbitrarily, but appropriately – what we shall understand by a language, we next set up a general calculus for all the objects that will prove to satisfy our definition, a calculus providing for all conceivable cases. This calculus is derived from our definition and in itself is founded, not on experience, but only on reasoning. We need not worry whether the calculus will fit given languages or given texts. The theory of linguistic structure must be so established as to have consistency and exhaustiveness of description built into the calculation, and this

requirement cannot be checked by seeing whether the theory really applies to all existing objects (such a check would be impracticable, both because of the number and size of the objects and because the theory must be applicable also to objects that do not yet exist), but only by reexamining the theoretical calculus to see whether it itself is consistent and exhaustive. In this matter the linguistic theoretician behaves like any other theoretician – like a mathematician, for example, who sets up this theory without regard to practical applications but whose theory, for precisely that reason, may be applied to cases that were not foreseen when it was established. An engineer can build a bridge simply by keeping to the mathematician's formulæ. Provided the mathematician's theory is consistent and exhaustive, and the engineer knows it sufficiently well and applies it appropriately, the bridge will hold.

A work of calculation such as we have here programmed in outline is obviously very comprehensive and very difficult. No such work was ever undertaken by classical linguistics, which did not even see the possibility of it. Much effort has since been put into the task, and many attempts have been made to carry it out, but it is still a long way to the establishment of a typology in which the known everyday languages would be distributed and on the basis of which general empirical propositions could be asserted about the structure of everyday language and its internal functions. For that reason, whatever we have to say about these matters must bear a provisional stamp.

We may take it as given that every language has two, and only two, planes: the content plane and the expression plane. In the theory we must also reckon with the possibility that a language may turn out, at the first stage of the analysis, to three or more planes, but it would be absurd to suppose such a structure is found in any everyday language and we can therefore ignore the possibility in our present discussion. Each of the two planes, as we have seen, can be divided further and further, down to the minimal entities, which in this book we have called *elements* – content elements and expression elements. These elements are ordered in *categories*, and a linguistic typology must show what categories can appear in a language, what categories can appear together, permitting or implying one another's presence, and what categories cannot appear together avoiding one another or implying one another's absence. Just as, within the chain, we distinguished between reciprocal or unilateral implicational relations and nonimplicational relations, so here we can distinguish between *reciprocal* or *unilateral implicational correlations* and *nonimplicational correlations*.

Within each of the two planes of a language, the most comprehensive categories into which the elements fall are those that we shall here call categories of *basic* and *characterizing elements*. This division rests on the special kind of relation that we call *government*. Now, not every kind of government will come into consideration here, but only what we shall call *direction* – the kind of government that serves to establish a clause or a unit composed of clauses. In the clause *sa première femme aimait les fleurs* there is an example of government in the fact that *sa* and *première* are feminine (in contrast to masculine *son* and *premier*) because *femme* is feminine and requires gender agreement of its adjectives; but in this instance the government is not a direction since it does not establish a clause, but only part of a clause (such a part – in our example, *sa première femme* – is called a junction). On the other hand, in the clause *sa première femme était jolie* the government connecting the feminine noun *femme* and the feminine form *jolie* is a direction since by means of it (and of other similar governments) the clause is established as a formal linguistic unit.

Now by a *characterizing element* we shall understand either an element that can enter as governed into a direction or an element entering into paradigm with such an element, while a *basic element* will be an element having neither of these properties. In the content plane of language, the characterizing elements are the *flexional elements* (e.g. ‘feminine’ in the examples given above), and the basic elements are the *stem elements* (e.g. the elements that enter into lexical words like ‘ram’ and ‘ewe’). In the expression plane, the characterizing elements are *accents* and *modulations*. In Danish there are two accents, one manifested in pronunciation by a (relatively) strong stress and the other by a (relatively) weak stress. The two have mutual commutation (*kórset* ‘the cross’ and *korsét* ‘corset’, whose expressions differ only in accent, have different contents), and they have unilateral implicational relation: the presence of a (relatively) weak stress in an utterance presupposes the presence of a (relatively) strong stress, but not vice versa. Moreover, this implicational relation is a direction, sufficient in itself to establish an expression clause. In standard Danish there are also two modulations. One, manifested in pronunciation by a falling tone, is found in apodoses and concluding clauses; the other, manifested in pronunciation by a non-falling tone, is found in conditional and introductory clauses and serves in general to give notice that the utterance is not finished and that more is to be expected (and if that “more” does not appear, the modulation permits us to supply it). As is already clear from what has been said, these two modulations have a unilateral implicational relation, the presence of a non-falling tone in an utterance presupposes the

presence of a falling tone, but not vice versa. Moreover, this implicational relation is a direction, sufficient in itself to establish a unit of expression clauses.

Provisional results indicate that every language has in its content plane both flexional elements (in our sense of the term) and stem elements; even the so-called isolating languages have certain “particles” that enter as governed elements into direction. The expression plane, however, does not always display the corresponding distinction between accent and modulation on the one hand and basic elements on the other. In ordinary written Danish we have neither accent nor modulation in the usual sense. To be sure, we have punctuation marks, like the question mark, which “points ahead” to a following answer (which may, if required, be supplied), but such punctuation marks are not an absolute necessity. Many languages lack accents in pronunciation. French is an example. To be sure, it is possible to interchange stronger and weaker stress in pronouncing French, but is never possible to distinguish the content of two words by a different distribution of stress as in Danish *kórset* – *korsét*. Whether all languages have modulations is an open question, if only because classical descriptions of languages generally give no account of modulations.

Thus we can expect to find in the system of human language (so far as concerns everyday language, to which we are here restricting our discussion) a unilateral implicational correlation – in the expression plane, at any rate – between characterizing elements and basic elements: a language cannot have accent and modulation without having basic elements of expression (e.g. vowels and consonants), but it can have the latter without the former. In the content plane there may perhaps be a reciprocal implicational correlation between flexional elements and stem elements.

These two main categories, or *species*, of elements – characterizing and basic – can be further divided into *types*. It would be too complicated to explain here how this division into types is actually effected according to the theory, but in practice we can say that we divide the characterizing elements into those that can be used to characterize a whole utterance or such as enter into paradigm with these and, on the other hand, those that lack this property. The former are called *extense*, and the latter *intense*. In the expression plane, the modulations are the extense elements and the accents are the intense elements. In the content plane, the extense elements are those that are commonly called elements of verbal inflexion (tense and mood, for example), while the intense elements are those that are commonly called elements of nominal inflexion (case and number, for example). The basic elements are divided into *central* – those that are indispensable in a

minimal unit, or such as enter into paradigm with these – and *peripheral* – those that are dispensable in a minimal unit. In the expression plane, where minimal unit involved is the syllable, the indispensable elements are the *vowels* and the dispensable elements are the *consonants*. In the content plane we find corresponding “content syllables” (they can be called minimal syntagmata and in practice will often coincide with individual words), divisible into indispensable or *radical elements* (e.g. the elements that enter into ‘faith’ in the relation ‘un-faith-ful’) and *derivational elements* (e.g. the elements that enter into ‘un-’ and ‘-ful’).

After what has been said above, we can reckon with the possibility of a language’s having modulations without having accents (French, for example), but hardly the other way round. Thus we have here a unilateral implicational correlation. It is a fact that a language cannot have consonants without also having vowels, but this is simply a consequence of the way we have defined vowel and consonant. Instead of vowels and consonants, a language may have a third, undifferentiated category of expression elements. Thus, there are languages with “vowel harmony” (e.g. Finnish, Hungarian, and Turkish), in which one vocoid (vocalic sound) in a chain determines the choice of all the other vocoids in the chain. In such languages, then, the vocoids represent accents, since they can enter as governed elements into a direction. And from this it follows that the contoids (consonantal sounds) do not represent consonants in such languages, but rather an undifferentiated category of expression elements. As for the content plane, it seems to be true that a language can have verbal, without having nominal, flexional elements (this is probably the case with the true isolating type of language, like Classical Chinese), but perhaps not the other way round. And apparently a language can have radical, without having derivational, elements, but certainly not the other way round. Here, then, there seems to be a unilateral implicational correlation.

The types, in turn, can be further divided into *subtypes* according to their implicational relations. As for the basic elements, the central type can be divided into those that must necessarily be present in the central part of the syllable (or syntagma) and those that need not be. Correspondingly, the peripheral type can be divided into those that must necessarily be present in the peripheral part and those that need not be. In the expression plane, for example, consonants can be so divided into two categories, one of which is represented by sounds or letters that stand at the beginning of the syllable, the other by those that stand at the end – or one category may be represented by those that occupy the first place in a group, the other by those that occupy the second place. The

division of the flexional elements into subtypes is particularly important. They may be divided into elements that can establish a compound sentence (or elements entering into paradigm with these) and elements that cannot. Case, for example, can serve only to establish a single clause: case direction never extends over the boundary of the single clause. Other flexional categories, however, do admit direction extending over the clause boundary: in *les fleurs sont jolies, et elles sont fraîches* a direction between feminine plural forms is to be found in each of the two clauses; but the feminine plural *elles* in the second clause is chosen because of a feminine plural in the preceding clause, which it presupposes, and there thus arises still another feminine plural direction binding the two clauses together into a compound sentence. This illustrates the way in which the various “grammatical categories” (case, gender, tense, mood, etc.) are defined in the linguistic structure.

The division into subtypes seems not to be universal in languages, but it is very common so far as flexional elements deemed. With the help of the commutation test we can discover how many elements are found in each of the category in a given language and what numerical relationships govern the structure of human language in general in this respect. Characteristically, the number of the elements is always relatively small. After all, the possibility of forming a practically unlimited number of signs with a small number of elements is implied in the very essence and purpose of language. But within the individual categories that happen to be represented in any given language, the number varies quite considerably. There are languages with two vowels and languages with twelve, or perhaps more, and all the numbers in between. There are generally more consonants than vowels. The number of accents and modulations seems always to be quite small, while in the content plane the number of flexional elements may be relatively small, as in English or Danish, or quite large. There are languages with two cases and languages with over fifty. There are languages with two numbers (singular and plural), with three (like Lithuanian or Ancient Greek, which also have a dual), and with four or five (like certain languages of the Pacific, with their trial and quadral numbers). There are languages with two genders (like French and Danish, if we leave the pronoun out of account), with three (German), with four (as in the Danish pronoun *han* ‘he’, *hun* ‘she’, *den* ‘it [common gender]’, *det* ‘it [neuter gender]’), and with as many as sixteen (Bantu). The least developed categories within the flexional system seem to be person, comparison, and verbal mood (English has two moods: indicative *lives*, imperative-subjunctive-infinitive *live-*, Danish has three: indicative *lever*, imperative *lev*, subjunctive-infinitive

leve-, German and Latin, in contrast to English and Danish, distinguish between the subjunctive and the infinitive, and Greek adds to these a separate optative).

This little survey is, of course, both incomplete and provisional, intended merely to suggest the first results that we might hope to obtain from a typology of linguistic structure. What we are eventually to gain from it – and what could not be shown in these examples – is the establishment of general laws stating what categories a language *can* have, what categories it *must* have, and what categories are bound together in a general implicational relationship.

In analyzing a language, we must add to the registration of the elements a registration of their *variants*. Two kinds of variants are to be found: the *varieties*, which are bound to their environments in a reciprocal implicational relation, and the *variations*, which vary freely and have only nonimplicational relation with their environments. The significance of this distinction can perhaps be seen most easily if we consider the representation of the expression elements in handwriting. A letter may take on different shapes depending on the other letters it is connected with: these are varieties. But each variety may in turn take on different shapes each time it is written: these are variations. Correspondingly, as shown by physical analysis, there are varieties and variations of spoken sounds. From the point of view of linguistic structure this division into variants is universal, i.e., it can be performed by an advance calculation on any entities whatsoever: any entity has as many varieties as it has possibilities of relation, and each of these varieties can be further divided into as many as an infinite number of variations. The peculiar fact is that the division into varieties and the division into variations are exhausted by turns: in a continued analysis there first comes a point at which varieties can no longer be divided into varieties, but only into variations; then comes a point at which variations can no longer be divided into variations, but only into varieties, and so on. We can see this most easily by considering variants of content entities. If we take the content of a sign like the word *man*, we can divide it into varieties until, for example, we reach the category of the men who can stand in the place where I am now standing, and at that point we cannot make any further division into varieties. But we can divide these varieties further into variations, namely, the different men who can stand in this place. These cannot be further divided into variations, but they can be divided into varieties – into each of these men in different places. These can be divided into variations – into each of these men in each of these places at different points of time. These can be divided into varieties – as seen from different points of view. And these, again, can be divided into

variations – as seen by different people. A variation that cannot be divided into variations, but only into varieties, we call an *individual*. A variety that cannot be divided into varieties, but only in variations, we call a *localized* variety. As is evident from the example, there are individual and localized varieties of different degrees.

Types of Linguistic Usage

Classical linguistics particularly cultivated the study of sounds (*phonology*, *phonetics*) from both physiological and physical points of view, and one might expect that a *phonetic* typology of languages could be set up as part of a typology of linguistic usage for the expression plane of language. Its aim would be to establish that certain linguistic communities use certain sounds while certain other linguistic communities use certain other sounds to represent the expression elements of their languages. And of course, in view of what we have said earlier, it would be concerned not merely with individual sounds, but with classes of sounds, sound *categories*.

In fact, classical phonetics has categorized speech sounds from many different points of view. It has divided them into sounds pronounced with more open mouth – what we shall call *vocoids* – and sounds produced with more closed mouth—the *contoids*. Both these main categories can be further divided in many ways: for example, according to the movements of the speech organs used to produce them, and according to the place where the movements occur. In this way the *vocoids* have been divided, for example, into open and close, front and back, rounded (i.e. pronounced with rounding of the lips) and unrounded, nasal and non-nasal. Similarly, the *contoids* have been divided into such categories as labials, dentals, and velars; stops, fricatives, and affricates; nasals and non-nasals; voiced and voiceless; and so on.

Against such divisions, the linguistically legitimate objection has been made that such categories can never have any fixed boundaries, whether from the physical (physiological) or from the linguistic point of view. From the physical (physiological) point of view, only continuous transitions are found between such categories. Between open and close sounds any number of half-open or half-close can be added. Between front and back *vocoids* can be found central *vocoids* and all sorts of other nuances. A sound can be strongly or weakly, wholly or partly nasalized, voiced, or rounded, and so on. Neither from the linguistic point of view are there any universally fixed boundaries, if we compare the expression elements of different languages and their usual phonetic representations. When we apply the commutation test to the expression elements in different languages, it becomes evident that there is no fixed relationship between the boundaries laid down in the phonetic sphere by the categories of a linguistic structure and the boundaries that can be established between the *phonetic* categories from a physical or physiological point of view. There is no simple coincidence between the vowel and

consonant categories and the physiologically defined categories of vocoids and contoids. In the Czech words *vlk* and *krk* (p. 38), *l* and *r* are vowels, but at the same time contoids (cf. also Eng. *little*). In the interjections of many languages, various contoids may be vowels (cf. *pst*, *hm*). Conversely, vocoids may be consonants (as, very frequently, *i* and *u*, which are then often represented orthographically by *y* and *w*). Since each language lays down its own boundaries, languages are not congruent from a phonetic point of view, and this will hold true as long as one continues the analysis. Because the number of expression elements differs from language to language, there clearly can be no fixed relationship between them and the sound categories that represent them. In French and English, voiced and voiceless *s* represent two different expression elements, as shown by the commutation test (Fr. *poison* ‘poison’ and *poisson* ‘fish’ differ in expression only in having voiced and voiceless *s* respectively; so also Eng. *zeal* and *seal*), but Danish makes no such distinction. In Danish, voiced *s* could be substituted anywhere for the customary voiceless *s* without any risk of confusion or misunderstanding, while that would be impossible in French or English. In French and Finnish, *t* and *d* are principally distinguished in pronunciation by the fact that *t* is unvoiced and *d* is voiced. In Danish, the principal distinction is that *t* is aspirated (followed by an *h*-sound) while *d* is not. Here again, different boundaries have been laid down within the phonetic sphere by the different languages, so that a Frenchman will take a Danish *d* for a *t*, and a Dane, unless he has special training, will be incapable of hearing any difference between *d* and *t* in French or in Finnish. On the other hand, no confusion or misunderstanding can possibly arise in French or in Finnish if an aspirated *t* is substituted for the unaspirated *t* that is usual in those languages, or in Danish if a voiced *d* is substituted for the usual voiceless *d*. Finnish makes no distinction between *k* and *g*, or between *p* and *b*; in both writing and pronunciation, only *k* and *p* are usually found. As a consequence, a Finn will be unable without special training to grasp the distinction that other languages make in this respect; and, on the other hand, it will be possible to substitute Dan. *g* and *b* or Fr. *g* and *b* for the usual Finn. *k* and *p* in any Finnish word without danger of confusion or misunderstanding. These are only individual examples; the situation is the same everywhere – the phonetic systems of languages are incongruent, and for that reason it is impossible to establish a universal phonetic system having any linguistic validity.

One might think to achieve better results by starting from some other matching of linguistic structure and linguistic usage than the one we have chosen. We have assumed that all conceivable (physically possible) sounds might represent, or be made to represent,

variants of a language's expression elements, even such sounds as are not customarily or deliberately used in the language under consideration. We have said, for example, that we *could* introduce a voiced *s* in Danish, or that in Finnish we *could* introduce an aspirated *t* instead of the unaspirated *t*, or *g* and *b* instead of *k* and *p*. It may be objected that while theoretically we could do this, in fact we do not. Danish *s*, when pronounced correctly, is *always* voiceless, Finnish *t* is *always* unaspirated and Finnish *k* and *p* are *always* *k* and *p*, never *g* and *b*.

This apparently very plausible objection, however, actually involves an extraordinarily rash assumption. How can we know these things? It is certainly impossible to examine everything that everybody has ever said in Danish or in Finnish, let alone what nobody has yet said but what will be said in those languages. For this reason, linguists in general wisely refrain from asserting that something *cannot be pronounced* in some particular way in a given language. Rather, they are content with saying in what ways (*among others*, understood) it *can* be pronounced. And yet there seems to be something tempting in the thought that within the whole continuum of possible sounds of human speech there might, in each linguistic usage, be certain no-man's-lands, zones unoccupied by any linguistic expression element. If so, it would not only be true that each language lays down its own boundaries in the phonetic domain, but it would also be true that each language selects its own areas therein. Thus, we might imagine a linguistic usage containing back and front vocoids, but no central vocoids, or unrounded vocoids, but no rounded ones. But whenever we undertake to investigate a linguistic usage in any considerable detail, we find ourselves disappointed in such expectations, and the thesis can hardly be maintained in so sharp a formulation.

We might perhaps try something else: following the English phonetician Daniel Jones, we might divide the sounds that appear as variants of a single speech sound (a single phoneme – i.e. a single phonetic representative of an expression element) into principal variants and secondary variants. In languages where front and back, but not central, vocoids are sounds of common occurrence, we could say that the different possible nuances of front and back vocoids were principal variants of their respective speech sounds (phonemes) while the central vocoids were subsidiary variants. Similarly, in languages (like English or Russian) in which unrounded front vocoids occur commonly, but rounded ones only in special cases, we could say that the former are principal variants, and the latter subsidiary variants, of their respective speech sounds (phonemes). But the very difficult question here is how one is to eliminate arbitrariness

and subjectivity from these decisions: whether something is “of common occurrence” or “frequent” is a matter of opinion.

It is unlikely, therefore, that we can construct a phonetic typology otherwise than by indicating what boundaries the linguistic structure can lay down in the phonetic domain. It is at least doubtful that we could declare certain zones to be no-man’s-lands within a given language and thus arrive at tenable general propositions about an implicational correlation between phonetic categories. An extraordinarily fascinating attempt in this direction has been made by the Russian-Czech linguist Roman Jakobson (now in the United States), who thinks it possible to show that children the world over learn the speech sounds in a fixed order – first certain sounds and then others – and that the child’s building up of the phonetic system of his language will be reversed if the faculty of speech is lost in aphasia resulting from a brain lesion, when the phonetic system will be broken down by the victim’s forgetting first the sounds that the child acquires last, and last the sounds that the child acquires first. Finally – and here we come to the decisive point – it is claimed that the resulting phonetic hierarchy will be reflected in the phonetic systems of linguistic usage, so that certain languages will have only the sounds that the child learns first and the aphasic forgets last, others will have, in addition, the sounds that the child learns next, and so on. With respect to children’s language and speech disturbances, advance theoretical objections can hardly be made against the hypothesis (but since it is derived solely from observation it must be either confirmed or invalidated by further observation, and it is impossible to predict what the outcome will be); in so far, however, as it is transferred to the study of the structure of phonetic systems, theoretical misgivings arise from the fact that it presupposes the existence of a phonetic no-man’s-land in a linguistic usage. Part of it can, in all likelihood, be maintained, even in respect of linguistic sound systems, namely the part involving an implicational relationship, not between sounds, but between linguistically determined phonetic boundaries. For example, although it probably could not be maintained that the existence of velar contoids in a language presupposes the existence of labial and dental contoids, it might possibly be maintained that the existence of, say, a linguistic boundary (i.e. a boundary separating representatives of two different expression elements) between front and back close rounded vowels (cf. Fr. *lu* and *loup*) presupposes the existence of a linguistic boundary between front and back half-close rounded vowels (cf. Fr. *feu* and *faux*). But this hypothesis, too, is based solely on observation and must therefore be confirmed or invalidated by observation; and here, too, predicting the outcome is risky.

In the study of meaning (*semantics*) we should expect to be able to arrive at a typology of linguistic usage for the content plane of language. This is for many reasons a more difficult task than phonetic typology, partly because semantics has been much less cultivated and partly because it embraces a far greater domain. The content of language is nothing less than the world surrounding us, and the minimal particular meanings of a word, the particular meanings that are individuals (cf. p. 114) are the *things* of the world: the lamp that stands here on my desk is a particular meaning of the word *lamp*; I myself am a particular meaning of the word *man*. But these *things* naturally organize themselves into many kinds of *categories*, and another difficulty is that one hardly knows what sort of a science is concerned with the establishment of these categories. With a certain justice one could say that it is all the sciences taken together. All sciences other than linguistics are actually theories of the linguistic content studied independently of the linguistic structure, just as physiological phonetics and physical acoustics are the study of the linguistic expression independently of the linguistic structure. In particular, we might think of psychology as being the science, if there is one, that could furnish such a categorization of experienced thing as would be suitable for mapping on the system of categories furnished by the linguistic structure. It is to be hoped that recent attempts of psychologists at a “phenomenological” description of the outside world as it is immediately experienced may lead to a fruitful collaboration with linguistics. It behooves us to maintain a more skeptical attitude toward the attempts – especially favored in former times – to found a linguistic semantics on the logic of ideas. For one thing, such an attempt runs in a vicious circle: conceptual logic is based on language (Aristotle’s logic, for example, would never have come to have the shape it has if it had not been thought in Greek); conceptual logic is always a language in disguise, and transforming or refining it seems to make no difference.

The prospects for a typology of meaning are therefore still poor. And we must realize in advance that it will face the same difficulties as the typology of sounds. Each language draws its own boundaries in the world of things and ideas. A sign in one language does not correspond to a sign in another language, as every translator knows. From the point of view of the world of things, just as of the world of sounds, there is no congruence between languages. And whether we can rightly find within the world of things or ideas certain zones that are subsidiary variants or no-man’s-lands in given languages is highly doubtful. Here, as in respect of phonetic typology, it will at best be a

matter of establishing implicational relationships, not between meanings, but between linguistically determined boundaries of meaning.

As the preceding considerations indicate, a typology of linguistic usage has crucial difficulties to contend with. The only kind of linguistic typology that seems realizable on an objective basis is, paradoxically enough, the one that classical linguistics never suspected to be possible: the typology of linguistic structure.

Linguistic Change

The great discovery that resulted from nineteenth-century research in genetic linguistics and that came to put its stamp on most of the linguistics of that period is that *languages change*. The different members of a language family may be considered as so many transformations of their common origin, the parent language, and in the history of the individual languages continual changes can also be observed and different stages can be distinguished. Because of the one-sided interest that classical linguistics had in usage, at the expense of structure, it was easy to exaggerate the importance of this discovery. For it is obvious that while linguistic structure is something relatively stable, which may well change in the course of time but which often remains constant over very long periods, linguistic usage is far more changeable. Words and other signs are ceaselessly arising and passing into obsolescence; pronunciations and meanings vary from place to place and shift from decade to decade, and indeed, under a more profound analysis, display a constant drift. For a science that saw nothing in language except signs, their pronunciations, and their meanings – and which, moreover, had celebrated its great triumphs in discoveries about linguistic change – this was bound to lead with almost fatal necessity to the abandonment of the concept of *linguistic state* (*état de langue*).

Meanwhile, sooner or later, this one-sided attitude was bound to come up against difficulties, and the very doctrine of language families and language change, when carried through, was to lead to a new and deeper understanding of the linguistic state. For it is clear that the parent language itself, known only as a set of formulæ designating the element-functions, is in itself a linguistic state and, what is more, a linguistic state in which all that is given is the linguistic structure. Even if we wished to go so far as not to admit any connexion between our Indo-European formulæ and any reality beyond the element-functions alone, it cannot be denied that when the formulæ are gathered together, as above on pp. 26-27, they constitute something which appears in all respects as the expression system of a language, a system of expression elements. To be sure, our formulæ – alias elements of the parent language – are each defined only as representing a certain function of elements between the genetically related languages. But we simply cannot avoid defining them by their functions *with one another* as well, ordering them in categories just like the expression elements of any other language, and dividing them into vowels, coefficients, consonants, accents, and so on. Moreover, it has proved to be not only unavoidable, but also highly fruitful, to establish rules governing the possible

combinations of vowels and coefficients in diphthongs, as well as a whole set of rules concerning word formation which are not merely – or essentially – concerned with the expression elements of the linguistic state involved, but first and foremost with its content elements, or at any rate its content units, and their possible combinations (cf. p. 29). Even if we do not venture to take the whole step of assigning real existence to the totality of our formulæ, and thus equating them with a linguistic state, we still cannot ignore the fact that these formulæ constitute a *system*.

Such a systematic view of the formulæ was first taken but the Swiss linguist Ferdinand de Saussure in a work dated 1879, which was to mark a turning point in the history of linguistics although the point of view that it represented was too strange to his contemporaries for it to meet with any general understanding. Indeed, even if the practical significance of Saussure's discoveries is now recognized by everyone working on genetic problems in Indo-European, it is perhaps only quite recently that we have become fully prepared to understand the theoretical scope of this work. Its distinctive feature is that, on the one hand, it treats the formulæ as a system and draws all the sequences that follow, while on the other hand it does not attribute any other kind of reality to them and thus does not consider them as prehistoric sounds with a certain pronunciation which by gradual change became the sounds of the individual Indo-European languages.

This point of view was bound to lead to practical advances in the elucidation of Indo-European itself. Precisely because Saussure treats the formulæ as a system and, moreover, as a system independent of any concrete phonetic definitions – in short, as a pure linguistic structure – he is led in this work to apply to the Indo-European parent language, the very stronghold of the theory of linguistic change, methods destined to set a pattern in the analysis of any *état de langue*, methods that can be taken as models of the way in which a linguistic structure should be analyzed. Saussure takes the system in and for itself and puts the question, How do I analyze it so as to obtain the simplest and most elegant explanation? In other words, What is the smallest number of formulæ or elements that I need to account for this entire mechanism?

And at this point Saussure came to do things with the Indo-European system that nobody had been able to do before, and thus to introduce a new method, a structural method into genetic linguistics.

Let us take an example. We have observed above (p. 30) that Indo-European has a vowel alternation $*e : *o : \emptyset$, which appears in diphthongs as $*e\grave{i} : *o\grave{i} : *I, *eu : ou : *u$, etc. In addition, one finds in Indo-European a different kind of alternation – or

something that might look like a different kind of alternation – namely, a shift between long vowel and *_A. This is seen, for example, in

OI <i>sthi-táh</i> ‘standing’	:	<i>ti-ṣṭhā-mi</i> ‘I stand’
Lat. <i>sta-tus</i>	:	<i>stā-re</i> ‘to stand’
IE root form * <i>st_A</i>	:	* <i>stā-</i>

or in

OI <i>di-táh</i> ‘given’	:	<i>dā-nam</i> ‘a gift’, <i>dá-dā-mi</i> ‘I give’
Lat. <i>da-tus</i>	:	<i>dō-num</i>
IE root form * <i>d_A</i>	:	* <i>dō-</i>

(We recall that OI *i* = Lat. *a* is IE *_A; see p. 17.)

Now Saussure realized that if the long vowel in these alternations were interpreted as a combination of short vowel with *_A, the two kinds of alternations, which had before looked entirely different, would become quite the same:

	* <i>e_i</i>	:	* <i>o_i</i>	:	* <i>i</i>
≠	* <i>e_u</i>	:	* <i>o_u</i>	:	* <i>u</i>
≠	* <i>e_A</i>	:	* <i>o_A</i>	:	* _A .

The row containing *_A thus becomes quite parallel with the rows containing **i*, **u*, or any other coefficient. All are special cases of the row **e* : **o* : \emptyset , and *_A is to be taken as a coefficient on a line with **i*, **u*, **r*, **l*, **n*, and **m*. (We may add that it was Saussure who introduced the term “coefficient,” and we may note at the same time that Saussure also introduced another term into linguistics, one that has since found high favor with linguists, namely, the term “phoneme” (cf. p. 118). He introduced it, for lack of a better word, as designation for the expression elements of language in order to avoid confusion with the “sounds” of linguistic usage – that is, to designate the purely “algebraic entities” of his theory. By an irony of fate, Saussure’s theory was so basically misunderstood by his contemporaries and by many who came after them that the term is now generally used as a synonym for “linguistic sound” – precisely what Saussure was trying to avoid.)

Perhaps this discovery of Saussure’s may not seem so very impressive. If so, the reason is that in the foregoing presentation of it the material has been organized for the

reader, and thereby the solution has been made fairly obvious. If we consider the commonly held notions of that time, however, the effect must have been quite surprising, if only because, as we have seen (p. 29), the short vowels $*e$ and $*o$ can also alternate with the long vowels $*\bar{e}$ and $*\bar{o}$ (cf., for example, the Greek nominative *rhétōr-a* ‘speaker’, with $*\bar{o}$, and the accusative *rhétor-a*, with $*o$) so that the long vowel alternating with $*_A$ might be thought to be functionally on a line with the long vowels alternating with zero and not, as it turned out, with the corresponding short vowels.

This reinterpretation of the long vowels alternating with $*_A$ as short vowel + $*_A$ was possible, of course, only because the combination of short vowel + $*_A$ does not appear otherwise in Indo-European; yet it meant a sharp break with the previous method of reconstruction because a formula like Saussure’s $*_{oA}$ is founded, not on the element-functions between the Indo-European languages, but on an internal function within the parent language. If the evidence were limited to the element-functions between the Indo-European languages known in Saussure’s time, there would be no grounds for distinguishing between the \bar{o} in *dō-num* and the \bar{o} in *rhétōr*. If the \bar{o} in *dō-num*, but not the \bar{o} in *rhétōr*, can be interpreted as $*_{oA}$, it is not because of a function between different languages, but because of a function between constituents of a single linguistic state. What has happened here is that one algebraic entity has been equated with the product of two others; the operation recalls that of the chemist who analyzes water as a product of hydrogen and oxygen. This kind of operation is required in the analysis of any linguistic state if the simplest possible description is to be achieved. Paradoxically, the first to which this operation was applied in the history of linguistic science was the Indo-European parent language.

In a paper published at the same time as Saussure’s and independently of it, the Danish linguist Herman Møller entertained similar ideas and discovered a new and very important simplification as well: the opposition between the three long vowels $*\bar{e}$, $*\bar{o}$, and \bar{a} can be interpreted, not as Saussure had interpreted it, by using a single coefficient $*_A$, but by using three. Thus, instead of $*\bar{e} : *_A : *\bar{o} : *_A, *\bar{a} : *_A$, it is possible to write $*_{aA_1} : *_{A_1}, *_{aA_2} : *_{A_2} : *_{aA_3} : *_{A_3}$. Later, Møller found that certain short vowels – $*e, *o, *a$ – could be explained in analogous fashion as products of the same coefficients and the same vowel: $*e = *_{A_2}a, *o = *_{A_2}a, *a = *_{A_3}a$.

To understand what is essential and important in these reductions from a methodological point of view, we must realize that they constitute a kind of resolution of the Indo-European entities into algebraic or chemical products. This does not proceed

directly from comparison of the several Indo-European languages, but from further treatment of the results of such comparison, from an analysis of those results. *Later*, long after the analysis had been made, it was discovered that there is an Indo-European language that distinguishes between $*\bar{o}$ alternating with $*o$ and $*\bar{o}$ alternating with $*_A$, namely Hittite, where the Polish linguist Kuryłowicz was able to show that *h* sometimes corresponds to IE $*_A$. Moreover, Møller was able to confirm his theory by referring to Hamito-Semitic: that Hamito-Semitic has special consonants corresponding to the different Indo-European coefficients is, in fact, a cornerstone in his demonstration of the genetic relationship between Indo-European and Hamito-Semitic. These confirmations, obtained through consideration of previously unknown element-functions between the genetically related languages, are certainly of the greatest interest, especially in showing that the internal analysis of a linguistic structure like that of the Indo-European parent language bears strong realistic implications. It might be thought that this kind of analysis would lead us astray in a world of abstractions, but quite to the contrary, it puts us on the watch for element-functions remaining to be discovered. Through analysis of the linguistic state we have truly achieved a deeper understanding of the linguistic structure. On the other hand, these confirmations from Hittite and Hamito-Semitic are still merely confirmations, and the internal analysis of the element system of the Parent language could be pursued without reference to them.

In the wake of these discoveries of Saussure and Møller concerning the structure of the parent language, various others followed, some made by the same two scholars and some others. The analytical principle involved, according to which an attempt is made to reduce to a minimum the number of element-formulæ needed for each stage of development language – a method that has been followed in the study of all linguistic states since its first application to the Indo-European parent language – has contributed to the introduction of a completely new technique within Indo-European genetic and comparative linguistics. In the chapter on genetic relationship we kept to the classical picture, obtained by mechanical registration of the element-functions between languages, but the system of formulæ that we ended up with (pp. 26-27) has been significantly reduced by this new principle of analysis. It would carry us too far afield, however, to go into details of the extensive reductions that result, interesting as they are. Suffice it to mention that it has been possible to explain combinations of consonant + *h* as combinations of consonant + coefficient $*_A$, and to explain the long vowels alternating with $*e$ and $*o$, as well as the zero with which they alternate, and accentual oppositions

(circumflex and acute; see p. 21) as varieties conditioned by the location of the word accent and by the number of latent syllables. Analytical reductions of this kind lead to an extraordinarily simple system of Indo-European formulæ.

There is another way, too, in which the linguistic state can be seen to assert itself and claim its rights in the face of an all too one-sided view of linguistic change. The best way to show this, perhaps, is by considering linguistic change itself. If – to use an artificial example for simplicity's sake – we find a language in which *p* has become *f* at the beginning of words, but has remained *p* in other positions, there must have been a period in the history of the language when every word-initial *p* mechanically changed to *f*. This would be a *sound law* in the true sense, on a line with any other law prevailing in a society, e.g. a juridical law: from the moment the law came into effect till the moment it ceased to prevail, it was in force and implied the mechanical conversion into *f* of initial *p* in any word coming within the society's range, whether the word was inherited, borrowed, or newly formed. (This is precisely why loan words may deviate in their external shape from inherited words: they have been borrowed after some law has ceased to prevail.) But is such a sound law truly a law of change or a law of state? If a society has a law decreeing that under certain conditions every young man of a certain age shall become a soldier, can it later be said that all such young men became soldiers? Undoubtedly it can, but the change referred to in those terms has its cause in a state of affairs. So long as that state of affairs continues – so long as the law is in force – it would be incorrect to speak of any change. What one can speak of is a mechanical conversion of one element to another under given conditions. So long as the law is in force, every word-initial *p* is mechanically converted to an *f*. This is the same kind of conversion law as the law in chess that a pawn, when it reaches the other side of the board, may have its value converted to that of a queen. The “change” takes place, or may take place, every time we play chess, and it will, or may, take place whether we play chess tomorrow or a year from now. It is not a change: it is a state.

The reader who is not a professional linguist will perhaps be surprised that the author can speak so warmly about so simple a matter. The reason is that only in our times are linguists working themselves free of a one-sided view of change. In fact, much of what has been considered one-sidedly from the point of view of linguistic change is to be seen as static phenomena. We have shown above (p. 16) that a special element-function can be established for IE **m* at the end of a word: in that particular position, Greek, for example, has *-n* instead of *-m*. But actually this must not be set up as a distinct element-

function. If we first give an account, as we must, of the static laws prevailing in each language, then this special element-function is superfluous. The fact that in Greek “-*m* becomes -*n*” is not an historical change that Greek has undergone, but a rule of the game that holds in the Greek system: in the structure of the Greek language, *m* is an expression element that is partly defined by its inability to stand at the end of the word, and therefore, whenever it would fall in that forbidden position, it must be replaced by -*n*. This is no more wonderful than the fact that we have to convert our money every time we cross the border between two countries. We must do so because each species of coin can be used only in its own place, and this is no law of change. Thus, linguistic change presupposes linguistic state, and thus, genetic linguistics presupposes the theory of linguistic structure, or grammar, which in fact means typological linguistics. Only after we have completely described the linguistic structure in all the individual linguistic states and found a place in them for everything that can be placed there, are we justified in proceeding to compare them from a genetic point of view. In this way, genetic linguistics will be simplified and clarified to an undreamt-of degree.

But even if more weight is thus assigned to linguistic state and less to linguistic change, the fact remains that language does change. And the linguist must seek a causal explanation of this change. This disputed and quite unclarified question is considerably simplified if we hold fast to the distinction between linguistic structure and linguistic usage. There can be many and various reasons for changes in linguistic usage, i.e. for changes in pronunciation (sound laws in the narrower sense), changes in meaning, changes in signs (words). Man is a capricious and enigmatic creature, and here, man is at work. The only provisional statement to be made with any certainty is that the sound changes undergone by a language in the course of time may often be summed up in certain predominant *tendencies* which a given population may stick to for centuries with remarkable tenacity. Maurice Grammont has shown strikingly how such tendencies toward altering pronunciation in certain directions have been at work in the eastern Indo-European languages from the oldest to the most recent times, constantly seizing hold of new material susceptible to change. What lies behind these tendencies is another question. They may be habits of a psychological nature. We encounter such tendencies in everyday life whenever we hear a foreigner speaking our native language. His “speaking with an accent” means precisely that by virtue of his tendencies of pronunciation, his psychological habits, he displaces the sound system of the language in certain directions. These tendencies are acquired, of course, not innate. But the difficulties connected with learning

to pronounce a foreign language like a native show that these acquired habits are almost ineradicable, and one can therefore understand how they can be stubbornly maintained from generation to generation over long periods of time.

The reasons for changes in the linguistic structure, on the other hand, cannot lie in tendencies of the speakers. Linguistic structure, after all, is defined as something independent of them. We assume that linguistic structure follows its own laws and that a change in linguistic structure is not due to tendencies of the speakers, but to *dispositions* in the system that undergoes change. A given linguistic structure may be thought of as being predisposed to move in certain directions and not in others. And here it is linguistic typology, with what it has to say about categories that seek out and favor each other and categories that shun or avoid each other, which must identify the causes of linguistic change.

Languages of Different Degrees

Every science aims to provide a procedure by means of which objects of a given nature are to be described. This is always done by introducing a *language* by means of which the objects can be described: a set of terms is introduced – a terminology with accompanying definitions – and then the description is made by using these terms to form sentences concerning the object in question.

Linguistics aims to provide a procedure by means of which languages are to be described. This is done by introducing a language by means of which languages can be described. Such a language for describing languages is called a *metalanguage*, and the language described is called an *object language*. Owing to the universalism of everyday language, an everyday language can be used as metalanguage to describe itself as object language: it is possible, for example, to write a Danish grammar in Danish. In general, however, it will be necessary to change somewhat the linguistic *usage* of the everyday language by introducing a number of new signs, the so-called technical terms. It is also possible to replace the everyday language, wholly or in part, with a specially invented symbolic language suitable for use as a metalanguage to describe other languages. This is also done in linguistics, but we have not used that kind of metalanguage in this book.

Thus we see that there can be languages of different degrees, first-degree languages, and second-degree languages, or meta-languages. Theoretically, of course, we can continue the progression: a language that describes a metalanguage will be a third-degree language, or a second-degree metalanguage (also called a meta-metalanguage). And we shall see that this theoretical possibility is actually found.

The grammar, or description, of the individual *état de langue* is thus a first-degree metalanguage. Using the terms introduced at various points in the preceding chapters, we can now briefly sum up the procedure of grammar by saying that it starts from a chain or a text as object for analysis, argues from that to a unilateral presupposition between the chain (the text) and the paradigms (the chain presupposes the paradigms), and, finally, on that basis, supplies the paradigms.

Like all sciences (as distinguished from the everyday languages), grammar must define its own signs as far as possible. But any science (and this includes grammar), however many of its signs it defines, will be forced to end its set of definitions somewhere, with the result that the signs entering into its basically presupposed definition will not themselves be defined. Thus in every science there will remain some indefinables,

or basic concepts – terms that are not defined within the scientific language itself but that can be defined only by the introduction of another language, which will be a metalanguage with respect to the language involved. So also in grammar there will remain certain basic concepts, certain undefined terms that can be defined only by the introduction of a second-degree metalanguage. Since all the defined terms of any science must necessarily be defined by means of other terms taken from the same language, the defined terms will be fixed by their mutual relationship and can never be fixed (or even partially fixed) by their relationship to any other objects than terms within the same language. In making use of glosses from an everyday language, grammar will have the advantage over other sciences that, when it is applied to the everyday language that it itself is using, it will be able to define all the terms it has taken from the everyday language. But, as we have seen, it can define them only by their relationship to other terms within the same language, not (either in whole or in part) by their relationship to objects that are not themselves terms within the same language. Thus, the terms grammar cannot be fixed by any things in some reality outside grammar to which they might refer or of which they are made (e.g. the electromagnetic vibrations caused by ink-marks on paper, or the sound waves caused by movements in the throat and mouth). Those, on the other hand, can be arrived at in meta-grammar, where we analyze the terms of grammar and find that the things to which they refer are the physical phenomena or other phenomena that furnish the expression and content of the object language (the everyday language) and that the things they are made of are the same as the things that the terms of other languages (e.g. the everyday language) are made of. In other words, grammar, like any other language, must be divided into a content plane and an expression plane. What constitutes the content in grammar, the content plane of grammar, is the expression and content of the everyday language. In order to treat within its content plane the expression elements and the content elements of the everyday language, grammar is obliged to introduce terms for them. It may, for example, call one expression element *p*, another *m*, and so on; and it may call one content element ‘he’, another ‘she’, another, perhaps, ‘sheep’, and so on. But all these terms introduced by grammar for the elements of the everyday language will remain undefined basic concepts, to be defined only in meta-grammar through an analysis of the content plane of grammar. Thus phonetics, the science of the sounds of language, and semantics, the science of the meanings of language, belong in meta-grammar. Between meta-grammar and grammar (as between any metalanguage and its object language) there is a relation of unilateral presupposition: the metalanguage

presupposes the object language, and meta-grammar thus presupposes grammar. Formulated in other terms: sounds and meanings presuppose elements of expression and content; the representation presupposes what is represented.

But there is also another way in which we find languages of different degrees. If we make a grammatical analysis of a given text, provided the text is long enough (e.g. Danish literature supplemented by a large collection of recordings of spoken Danish), we shall find that within the whole text there can appear – or by extension of the text there can be produced – fragments (either textual components of large size, or individual words, or individual elements, or individual variants) which can be translated within the text into other fragments. One fragment, perhaps, is in Jutland dialect, another in Zealand dialect; and it turns out to be possible to translate the Jutlandic fragment into Zealandic, and the Zealandic into Jutlandic. Or one fragment is written language and another spoken language, and they too can be mutually translated. Or one fragment is everyday speech, another lecture style; one fragment is prose, another poetry – in all cases, mutual translation is possible. If we had a text including different national languages, like English and Danish, we should be faced with the same situation. And the moment we have a text recorded by two different persons, we again have the same possibility of translation. Every linguistic physiognomy ("voice," "handwriting") has its own distinctive stamp, in the linguistic usage at any rate, and a text can be translated from one physiognomy to another. If I have a number of students read aloud, or relate, or write down the same story, I obtain a number of translations of one and the same text into different physiognomies.

Each time we find such a translatability between two fragments of a text, we must take account of it. Whenever we encounter different national languages, local dialects, jargons, idioms, codes, styles, or physiognomies, we must analyze separately the text that each of them has yielded, because they may reveal differences in linguistic structure. Thus, in grammar, we have to provide ourselves with a system of labeling, so that one group of texts can be marked "English" and another "Danish," or one "prose" and another "poetry," or one "John Smith" and another "James Brown," and so on – divisions which obviously cross one another in many ways. But the labels themselves – "English," "Danish," "prose," "poetry," etc. – cannot be further treated within the language of our grammar.

For the remarkable thing is that these designations stand for notions that are generally of sacral character, involving the use of a given linguistic usage or of a given linguistic structure as expression for a content consisting of certain elements that lie

outside the language. Thus the Danish language is expression for a content that is the Danish nation, the family, and the home; and in the same way, different styles are expressions or symbols, for contents consisting of certain elements that lie outside the styles. So we are again dealing with an expression and a content – with a language – but this time with a language whose expression plane is itself a language, with its own content and expression. Thus we must add to our grammar a new grammar, which treats all these “labels” (or connotators, as we shall call them) as content for a given language as expression (and we shall call this content-expression function a connotation). So here again we have to do with grammars of various degrees, but in a different sense from before. And it may be appropriate to continue along this line: in a third-degree grammar we might show what geographical, historical, social, and psychological elements compose the content of the name of a language – “Danish,” for example – considered as expression for such content. Everything from national down to personal characteristic finds a natural place within the sphere of linguistics.

Linguistic genetics and linguistic typology are, of course, no different languages from grammar; they are merely extensions of its domain. Each, as we have seen, treats certain functions between the *états de langue*. Using the terms we have introduced above, we can summarize by saying that linguistic genetics supplies a parent language, and linguistic typology supplies a type, behind the individual languages. There is a correlation of unilateral presupposition between the individual languages and the parent language: the individual languages presuppose the parent language, but no vice versa. And the function between individual languages and parent language is a correlation, not a relation: they are not present together, but alternate. Between the individual languages and the linguistic type there is a relation of unilateral presupposition: the individual languages presuppose the linguistic type, but not vice versa. And the function between individual language and linguistic type is a relation not a correlation: individual language and linguistic type are present together; if we are given an individual language, then in it, *eo ipso*, we are given its type. At this point we have finally reached a *formal definition* of the two kinds of linguistic relationship: genetic relationship rests on a continuation, i.e. a correlation of unilateral presupposition between languages; typological relationship rests on type-realization, i.e. a relation of unilateral presupposition between languages.

Index

- * A, 17, 27, 87, 124-28
- Abbreviation of signs, 56-59
- Accent, 12, 13, 20-21, 29, 95, 108-10, 112, 128
- Acoustics, 120
- Afghan, 72
- Afrikaans, 69
- Agglutinating type, 92-93
- Ainu, 80
- Akkadian, 73
- Albanian, 65, 72
- Aleut, 81
- Altaic, 77-78, 80, 92, 93
- Alternation, 29-30, 124-28
- American languages, 21, 39, 81, 93 Amharic, 76
- Analogical formation, 51-56
- Analysis, 8, 97, 99-100, 102-3, 113, 126, 133
 - Anglo-Frisian, 69
 - Aphasia, 119
 - Arabic, 73, 76
 - Aramaic, 73-76
 - Armenian, 11, 72, 93-94; cited, 11, 12, 13, 14, 17, 18, 19, 20, 82, 93
 - Australian languages, 80
 - Austic, 78, 79
 - Austro-Asiatic, 78
 - Avestan, 72
- Baltic, 53, 71
 - Baltic Finnish, 77
- Bantu, 76, 79, 80, 112
- Basic element, 107-9, 111
- Basque, 81
- Belorussian, 72
- “Bokmål,” Norwegian, 68
- Borrowing, 59-65. *See also* Loan words
- Breton, 69
- Brythonic, 69
- Bulgarian, 71
- Bushman, 81
- Canaanite, 73
- Case, 95, 109, 111-12
- Catalan, 70
- Category, 33, 35-36, 56, 57, 95-96, 97-98, 100, 107, 111, 112, 115-16, 120, 131
 - Category-function, 95
 - Caucasian languages, 80
 - Celtic, 11, 64, 69
 - Central element, 109-11
 - Chain, 32, 97, 133
 - Chaldean Pehlevi, 72

- Change, linguistic, 96, 122-31
- Characterizing element, 107-9
- Cheremiss, 77
- Children's language, 119
- Chinese, 69, 76, 78, 92, 94, 110
- Chuvash, 77
- Class, 8
- Classical linguistics, 3, 6-7, 26, 85, 87, 91, 94, 96, 106, 115, 121, 122
- Clause, 98, 99
- Code, 41, 43
- Coefficient, 27, 125
- Common Germanic, 22
- Common Romance, 22
- Commutation, 100-101, 111, 116
- Comparison, 112
- Connotator, 136
- Consonant, 26, 35, 95, 98-99, 110-12, 116
- Constant, 40
- Contamination, 54
- Content, 32, 55, 88, 95, 99, 101, 102-8, 120, 134, 136
 - Content definition, 103
 - Content element, 32, 100, 102-4, 106
 - Continuation, 30, 137
 - Contoid, 110, 115-16, 120
 - Coptic, 73, 80
 - Cornish, 69
- Correlation, 33; reciprocal implicational, 107; unilateral implicational, 107, 110, 136-37; nonimplicational, 107
 - Costumes, 101
- Counter-examples, 30-31, 45, 46, 54, 56, 64, 66
 - Covered element, 47-48
 - Cuneiform, 11, 73
 - Cushitic, 73
 - Czech, 38, 72; cited. 38, 116
- Dalmatian, 71
- Danish, 25, 59, 65, 66, 108, 112, 116-18; cited, 18, 23, 24, 46, 49, 54, 56, 58, 59, 61, 66, 88
 - Definition, 30, 103, 133
 - Degrees of languages, 132-37
 - Dependence. *See* Function
 - Derivational element, 110
 - Description, 8, 121, 132-33
 - Diphthong, 27
 - Direction, 107
 - Dispositions, 131
 - Dissimilation, 48-49, 66
 - Division. *See* Analysis
 - Dravidian, 80
 - Dutch, 69
- EE. *See* Expression element
- EE-function. *See* Element-function
- Egyptian, 73, 80
- Element, 97-98, 106. *See also* Basic e.; Central e.; Characterizing e.; Content e.; Covered e.; Derivational e.; Expression e.; Extense e.; Flexional e.; Intense e.; Peripheral e.; Radical e.; Stem e.; Uncovered e.
- Element-function, 13, 15-26, 45, 46, 48, 61, 64, 80, 84, 86, 87, 95, 122-23
- Elements, system of, 35-37, 39; action
 - of, 45-51; types of, 107-8; species 109;

- subtypes of, 110-11
- English, 11, 22, 59, 65, 69, 71, 76 86 94, 112, 116-17, 119; cited, 23, 24, 32, 46, 50, 53, 54, 55, 56, 58, 59 60, 61, 64, 66
- Eskimo, 81
- Estonian, 67, 77
- Etat de langue*, 8, 9, 97, 122-24, 127-30
133, 136
- Ethiopic, 76
- Etruscan, 81
- Etymology, 90; popular, 54, 60
- Evenki, 78
- Everyday language, 101-2, 104-5, 106, 109, 132-33
- Expression, 4, 95, 99, 101, 102, 108-12,
134, 136
- Expression definition, 103 Expression element, 13, 15-17, 31, 32, 33, 37, 84, 100, 101-2, 109
Extense element, 109

- Faeroese, 70
- Family, language, 9, 21-22, 70-81, 122 Feeling for a language. *See Sprachgefühl*
- Field, 47
- Finnish, 39, 65, 77, 110, 117-18; cited, 63
- Finno-Mordvin, 77
- Finno-Permian, 77
- Finno-Ugrian, 76-77
- Flemish, 71, 72
- Flexional element, 107-12
- Folk costumes, 101
- Foreign words, 60. *See also* Loan words
- Formula, functional, 15-30, 81
- French, 22, 39, 46, 61, 64, 65,
94, 109, 110, 112, 116-17; cited
49, 54, 56, 57, 58, 59, 116, 120
- Frisian, 69
- Function, 8-9, 13, 15, 32, 33. *See also* category-function; Element-function
- Functional transformation of signs, 45-59

- Gaelic, 69
- Galician, 70; cited, 49
- Gallo-Romance, 70
- Games, 42-43, 99, 101
- Gaulish, 69
- Gender, 95, 111-13
- Generalization, 62
- Genetic linguistics, 6-7, 85, 89-90, 96,
122, 124, 128, 130, 136
- Genetic relationship, 6-7, 9, 10-31,
78-79, 82-83, 94, 95, 127, 130, 137
- Georgian, 38-39; cited, 82
- German, 11, 22, 24, 25, 59, 61, 69, 71,
76, 87, 112; cited, 23, 49, 54, 55, 56, 58, 59
- Germanic, 22, 24, 46, 53, 54, 64, 65, 68-69, 72; sound shift in, 19-20, 24, 87
- Gothic, 11, 22, 68; cited, 11, 12, 13, 14,
18, 19, 20, 23, 24, 45, 51-52, 53, 54, 56, 88
- Government, 107
- Grammar, 8, 130, 133

- Grammont, Maurice (1886-1946), 51, 130
 Great Russian, 72
 Greek, 11,63,64,67,71, 112, 121, 129- 30; cited, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 28, 29, 45, 50, 51-52, 82, 88, 126
 Greenlandic, 92-93, 94; cited, 93
- Hamitic, 73
 Hamito-Semitic, 73-76, 79, 80, 127
 Haplology, 50
 Hebrew, 73-76
 Hellenic, 71
 Hieroglyphics, 73
 High German, 24, 25, 69, 76, 87; sound shift in, 24, 87; cited, 24
 Hispano-Romance, 70
 History, 4-5
 Hittite, 11, 73,127; cited, 14,18,51-52
 Homburger, Lilius, 79
 Hottentot, 81
 Humboldt, Wilhelm von (1767-1835), 93
 Hungarian, 76-77, 112; cited, 63
 Hypotheses, 85, 87
- Icelandic, 64, 68; cited, 24, 63
 Icelandic sagas: language of, 11
 Indefinables, 133
 Indic, 11, 72
 Individual, 114
 Indo-European, 10, 68-73, 79, 80, 92, 93
 Indo-European parent language, 26- 27, 85-89, 123-24, 127-28
 Inflecting type, 92, 93
 Insular Brythonic, 69
 Intense element, 109
 Iranian, 72
 Irish, 11, 69; cited, 11, 12, 13, 14, 17, 18, 19, 20, 45, 51, 64
 Isolating type, 91-92, 93, 94, 108
 Italian, 22, 70, 71; cited, 22, 46, 49, 50, 53, 56
 Italic, 22, 69-71
 Italo-Sardinian, 70
- Jakobson, Roman, 119
 Japanese, 80
 Jones, Daniel, 118
 Yugoslav, 71
 Junction, 107
- Karelian, 77
 Khoisan, 81
Kierkegaard, Søren (1813-55), 104
 Komi, 77
 Korean, 80
Kuhn, Ernst (1846-1920), 79
 Kurdish, 72
Kurylowicz, Jerzy, 127

"Landsmil," Norwegian, 68-69

Language, 36-37, 101, 104, 132. *See also* Everyday language

Lapp, 77

Latency, 98, 128

Latin, 11, 22, 47, 59, 60, 63, 64, 65, 70, 84, 86, 112; cited, 11, 12, 13, 14, 17, 18, 19, 20, 32, 46, 48, 49, 50, 51-52, 53, 54, 56, 57, 63, 64, 66, 88, 92, 124-26

Law, 128-29

Lepontic, 70

Letter changes, 87

Letters, 13, 40-41, 43, 86-87, 113

Lettish, 71

Leveling, 51-54

Libyco-Berber, 73

Linguistics, 3-7, 96. *See also* Classical linguistics; Genetic linguistics

Lithuanian, 11, 72, 112; cited, 11, 12, 13, 14, 15, 17, 18, 19, 21, 28, 45, 51, 52, 53, 54, 56, 57, 58, 82

Loan words, 59-65, 82, 129

Localized variety, 114

Logic, 4-5, 91, 121

Low German, 24, 61, 65, 69

Lusatian, 71

Lycian, 73

Macedonian, 71

Malayo-Polynesian, 78, 79

Manchu, 78

Manx, 69

Mari, 77

Meaning, 87-88, 95, 100, 120-21, 130, 134

Messapic, 73

Meta-grammar, 134

Metalanguage, 132-34

Metathesis, 50

Middle English: cited, 18

Middle High German: cited, 50

Middle Indic, 72

Middle Irish: cited, 18

Minimal sign, 88, 91

Minimal syntagma, 110

Moabite, 73

Modulation, 95, 108-9, 112

Møller, Herman (1850-1923). 79.

126-27

Mongol, 77-78

Mon-Khmer, 78, 79

Mood, 109, 111

Mordvin, 77

Munda, 78, 79

Neologism, 66-67

Nilotic, 79

Norse. *See* Scandinavian

Norwegian, 59, 68-69; cited, 54

Nostratic, 80

Number, 95, 109, 112

- Nynorsk, 69
- Ob-Ugrian, 76
- Old Canaanite, 73
- Old Church Slavonic, 11, 71; cited, 11, 12, 13, 14, 15, 17, 19, 20, 45, 51, 53, 54, 57
- Old Egyptian, 80
- Old English, 11; cited, 23, 24
- Old French, 60; cited, 56, 66
- Old High German, 11; cited, 18, 23, 46
- Old Indic, 11, 72; cited, 11, 12, 13, 14, 15, 17, 18, 19, 20, 45, 48, 50, 51, 52, 53, 88, 124-25
- Old Norse, 11, 22, 70; cited, 12, 14, 15, 17, 18, 20, 23, 51-52
- Old Persian, 72
- Old Prussian, 71
- Old Saxon, 24
- Old Turkic, 77
- Old Welsh: cited, 18
- Operational definition, 30
- Orkhón inscriptions, 77
- Oscan, 70
- Osmanli, 77
- Ossetic, 72
- Ostyak, 76
- Palæo-Asiatic languages, 80
- Pāli, 72
- Paradigm, 33, 51, 53-54, 55, 133
- Parent language, 82-90, 122-23, 136
- Pedersen, Holger (1867-1953), 80
- Pehlevi, 72
- Peripheral element, 109-11
- Permian, 77
- Permutation, 91, 92
- Persian, 72
- Person, 112
- Phenomenology, 121
- Phoenician, 73
- Phoneme, 118-19, 125
- Phonetics, 4, 115-20, 134
- Phrygian, 71
- Physics, 4-5, 47, 113, 115, 120, 134
- Physiognomy, linguistic, 135
- Physiology, 4-5, 47, 48, 115, 120,
- Polabian, 72
- Polish, 72; cited, 18
- Polysynthetic type, 92-94
- Popular etymology, 54, 60
- Portuguese, 70; cited, 49
- Prakrit, 72
- Prediction, 89
- Pronunciation; change of, 130-31
- Proportional formation, 55-56
- Proto-Germanic, 22
- Proto-Romance, 22
- Provençal, 70; cited, 66
- Provençal, 70; cited, 66

- Psychology, 4-5, 44, 46, 55, 121, 131
 Pushtu, 72
- Radical element, 110
 Rask, Rasmus (1787-1832), 19, 23, 87
 Reconstruction, 83
- Relation, 32-33, 35; implicational, 97; nonimplicational, 97-98; reciprocal implicational, 98-99; unilateral implicational, 98-99, 134, 136-37
- Relationship. *See* Function; Genetic r.; Typological r.
 Representation of elements, 40-41, 48, 86-87, 113, 117, 134
 Rhæto-Romance, 71
 "Riksmål," Norwegian, 68-69
 Romance, 22, 46, 63, 70-71, 72
 Romany, 72
 Rumanian, 22. 71; cited, 22
 Runic inscriptions, 68
 Russian, 59, 72, 119; cited, 18 38, 58 59
- Sakian, 72
 Samnite, 70
 Samnorsk, 69
 Samoyed, 76
 Sanskrit, 72. *See also* Old Indic
 Santālī, 78
 Sapir, Edward (1884-1939), 93
 Sardinian, 70
 Saussure, Ferdinand de (1857-1915), 123-27
 Scandinavian, 11, 22, 24, 25, 60, 65, 68-69
 Schlegel, Friedrich von (1772-1829), 93
 Science, 5, 85, 96, 120, 132-33
 Scottish Gaelic, 69
 Semantics, 120-21, 134. *See also* Meaning
 Semitic, 64, 73-76
 Serbo-Croatian, 71
 Siamese, 78
 Sicilian, 70
- Sign, 31, 32, 34-35, 37-40, 45, 88, 91, 102-3, 132-33
 Sign formation, 34, 42, 45-67
 Signs: functional transformation of, 45-59; action of, 51-56; abbreviation of, 56-59; introduction of new, 59-67
- Sino-Austic, 6, 78, 79, 80
 Sino-Tibetan, 79
 Slavic, 53, 63, 71-72
 Slovak, 72
 Slovene, 71
 Sociology, 4-5
 Sogdian, 72
 Sorbian, 71
 Sound law, 87, 128-30
 Sound shift: Germanic, 19-20, 24, 87; High German, 24, 87
 Sound-symbolism, 43
 Spanish, 22, 70; cited, 22, 46, 49, 53, 58
 Species of elements, 109
 Speech sounds, 4, 13, 41, 43, 87, 95, 113, 115, 118-19, 125, 134
Sprachgefühl, 44, 45, 47, 54, 55, 60, 61

- State, linguistic. *See Etat de langue*
 Stem element, 107-108
 Stress, 47
 Structural method, 124
 Structure, linguistic, 7, 31, 32-44, 53, 57, 58, 60-61, 91, 95, 96, 122, 130-31; types of, 97-114
 Subfamily, 21-26
 Subtypes of elements, 110-11
 Sudanese languages, 79-80, 81
 Sumerian, 81
 Swedish, 25, 54, 59, 68
 Syllable, 34-37, 47-48, 98
 Syntagma, 110, 111
 Synthesis, 8
 System, 123
- Tabu, 65-66
 Tarski, Alfred, 104-5
 Tendencies, 130-31
 Tense, 109, 111
 Text, 97, 133
 Thai, 78
 Thai-Chinese, 78, 79, 93
 Things, 120
 Thomsen, Vilhelm (1842-1927), 65, 77
 Thracian, 73
 Tibeto-Burman, 78, 79
 Tokharian, 11, 73; cited, 11, 12, 13
 Transformation of signs, functional, 45-59
 Translatability, 135
 Tungus, 78
 Turkic, 77
 Turkish, 63, 77, 94, 110; cited, 92
 Typological relationship, 9, 91-96, 137
 Typology of languages, 9, 91-96, 121 130-31, 136
- Udmúrt, 77
 Ugrian, 76
 Ukranian, 72
 Umbrian, 70
 Umbro-Samnite, 70
 Uncovered element, 47-48
 Uniforms, 101
 Units, 95-96
 Universal analysis, 113
 Uralic, 76-77, 80, 92, 93
 Usage, linguistic, 31, 32-44, 54, 57 5g 60-61, 91,95,96, 115-21, 122, 130-31, 132
- Variable, 40
 Variant, 58, 113-14, 117-19
 Variation, 113-14
 Variety, 113-14, 128
 Vedic, 72. *See also* Old Indic
 Venetie, 73
 Verner, Karl (1846-96), 20, 24
 Vietnamese, 78, 79

Vocoid, 110, 115-16, 118-19
Vogul, 76
Votyak, 77
Vowel, 27, 35, 95, 98-100, 110-12, 116
Vowel alternation, 29-30, 124-28
Vowel harmony, 110
Vulgar Latin, 22, 46

Welsh, 11, 71; cited, 14, 18,64
Wendish, 71
White Russian, 72
Word, 31, 32, 91, 110
Word change, 130
Word formation, 28-30
Word structure: typology of, 91-93
Wulff, Kurt (1881-1939), 79

Yagnobi, 72
Yakut, 77
Yiddish, 76

Zarathustra, 72
Zyrian, 77

φ , 15
*, 16