



Cadernos de Tradução
do Instituto de Letras

Número 43, jul/dez — 2018

A Linguagem através dos Estudos do Léxico e da Lexicografia

**Sandra Dias Loguercio
Ana Eliza Pereira Bocorny
(orgs.)**



UFRGS

UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

INSTITUTO DE LETRAS — UFRGS

Direção

Sérgio de Moura Menuzzi (Diretor)

Beatriz Cerisara Gil (Vice-Diretora)

COMISSÃO EDITORIAL

Andrei Cunha

Sandra Dias Loguercio

Rozane Rebechi

ORGANIZAÇÃO DESTE NÚMERO

Sandra Dias Loguercio

Ana Eliza Pereira Bocorny

Universidade Federal do Rio Grande do Sul

Instituto de Letras

Av. Bento Gonçalves, 9500 — CEP 91540-000 — Porto Alegre — RS

<http://www.ufrgs.br/letras/index.html>

Como citar este periódico (ABNT):

Cadernos de Tradução, Porto Alegre, n. 43, jul/dez 2018.

Cadernos de Tradução

**A Linguagem através dos
Estudos do Léxico e
da Lexicografia**

**Sandra Dias Loguercio
Ana Eliza Pereira Bocorny
(orgs.)**

Sumário

A LINGUAGEM ATRAVÉS DOS ESTUDOS DO LÉXICO E DA LEXICOGRAFIA

Apresentação / 6

Sandra Dias Loguercio

Métodos e técnicas para usar a Internet diretamente como *corpus*: o caso dos Dicionários on-line de Espanhol Valladolid-UVa / 10

Sven Tarp; Pedro A. Fuertes-Oliveira

Tradução: Luísa Rabaldo

Revisão: Ana Eliza Pereira Bocorny; Sandra Loguercio

Dicionários sensíveis ao contexto e integrados a assistentes de escrita em L2: novos desafios para a Lexicografia / 33

Sven Tarp; Kasper Fisker; Peter Sepstrup

Tradução: Guilherme S. de Oliveira

Supervisão e revisão: Márcia Moura da Silva

Dicionários e exemplos de apoio à produção linguística / 63

Ana Frankenberg-Garcia

Tradução: Luísa Rabaldo

Revisão: Rozane R. Rebechi

Construir para desconstruir e voltar a construir: elaboração colaborativa de um dicionário escolar de ciências / 87

Rosa Estopà

Tradução: Tiele Kowarlevski

Supervisão: Líliam Ramos

Revisão técnica: Cleci Bevilacqua

Informações culturais em um dicionário bilíngue de aprendizagem / 113

Michèle Fourment-Berni Canani

Tradução: Samanta Siqueira

Supervisão e revisão: Patrícia Reuillard

Revisão técnica: Sandra Loguercio

Características distributivas do Espanhol de Cuba a partir da análise de dois estudos lingüísticos / 126

Lucía Marconi; Leonel Ruiz Miyares; Paola Cutugno

Tradução: Cyrano da Rosa Silva

Supervisão e revisão: Cleci Regina Bevilacqua

Os verbos de constatação no texto científico / 149

Francis Grossmann

Tradução: Alex de Cássio da Silva

Supervisão e revisão: Patrícia Reuillard

Revisão técnica: Sandra Loguercio

A fraseologia transdisciplinar dos textos científicos: das colocações às rotinas retóricas / 170

Agnès Tutin

Tradução: Anna Carolina Teles

Supervisão e revisão: Patrícia Reuillard

Revisão técnica: Sandra Loguercio

APRESENTAÇÃO

*“Quem tem mais palavras ou que sabe melhor manusear
as poucas que tiver **está** em boa condição de jogo”*

Elisa Lucinda

Estudos dedicados à descrição lexical ou à sua faceta aplicada, a Lexicografia, não prescindem mais dos recursos que as evoluções tecnológicas do final do século XX, marcadas, entre outras, pelas tecnologias da informação e da comunicação (as TICs), nos propiciam. Além da possibilidade de trabalhar com grandes *corpora* textuais, cujos conteúdos podem ser rapidamente mapeados, calculados, organizados de diferentes formas com a ajuda de programas de processamento automático da linguagem, temos acesso a uma infraestrutura comunicacional que permite a interação em rede, multiplicando a capacidade de troca de informação e do trabalho colaborativo. A exemplo do que vemos com a *Wikipédia* e com dicionários *online*, como o *Reverso*, usuários passam a poder contribuir com os recursos de consulta, acrescentando-os de todo tipo de informações. Muda, assim, nossa maneira de olhar para as línguas e para a produção de linguagem de modo geral e mudam os recursos dos quais dispomos para a comunicação, que nunca foi tão facilitada e enriquecida de possibilidades.

Por outro lado, para usufruir desses avanços e fazer parte do “jogo”, há que se saber “manusear as palavras”. Os processos de interação que se realizam também hoje, e muitas vezes sobretudo, em ambiente virtual, feitos *grosso modo* de leitura e produção de texto, não deixaram de evidenciar as questões que nos fazemos desde sempre quando a compreensão ou a capacidade de expressão falham: *o que é isso? O que quer dizer? Ou, por outro lado, como se diz? Como se traduz? Como posso ser mais preciso?* Perguntas, na realidade, que exigem estudos sofisticados para serem bem respondidas, uma vez que expõem a conexão da linguagem com o mundo e com a própria interlocução, intensificando-se com a interculturalidade em todas suas dimensões, seja entre falantes de uma mesma língua seja entre falantes de línguas diversas. E para obter respostas a essas questões, é à *palavra* que, enquanto leitores e produtores de texto, nos dirigimos. Por mais ambígua que seja – do que afinal estamos falando ao tratar de *palavra*: de *em*, de *em condição de*, de *boa condição*, de *em condição de jogo*? Qual *jogo*? –, é pela palavra que

acessamos as respostas que buscamos sobre o mundo e sobre a própria linguagem através de textos.

Os estudos do léxico e de suas ferramentas de consulta, mais conhecidas na forma de dicionários, fazem da *palavra* sua unidade privilegiada para abordar a linguagem e suas diversas situações de produção. A linguagem produzida pelos mais distintos contextos da atividade humana é evidenciada pelos estudos de Terminologia; a linguagem produzida pelas práticas tradutórias, que criam encontros entre culturas, estimula e, ao mesmo tempo, coloca à prova tanto a produção científica voltada aos estudos do léxico quanto a produção lexicográfica; a linguagem vista da perspectiva das variadas situações de letramento é objeto de reflexão da Lexicologia e do que se convencionou chamar de Lexicografia Pedagógica, dedicada à análise e elaboração de instrumentos de consulta lexical que visam favorecer a aprendizagem linguística. Esses estudos já têm uma história em nosso Instituto de Letras, consagrada pela criação da linha de pesquisa de Pós-Graduação *Lexicografia, Terminologia e Tradução: Relações Textuais*, e estão em diálogo constante com a produção científica produzida em diferentes contextos nacionais e internacionais. É com o intuito de compartilhar leituras desse campo de pesquisa e ampliar o número de interlocutores que convidamos os autores aqui apresentados a disponibilizarem seus trabalhos via tradução.

Iniciamos com o artigo de Sven Tarp e Pedro A. Fuertes-Oliveira, que precisam as diferentes situações em que o lexicógrafo pode explorar dados empíricos obtidos da Internet e propõem métodos para qualificar e tornar mais eficientes essas práticas. Isso passa, entre outras etapas, pela busca de informações sobre as necessidades lexicográficas do público-alvo de um dicionário que se quer elaborar, pela seleção e preparação das informações lexicográficas (o uso da Internet como *corpus*) – cuja discussão é ilustrada pelo projeto em andamento dos Dicionários On-line de Espanhol Valladolid-Uva – e pela avaliação do dicionário do ponto de vista do consulente.

Sven Tarp também contribui, em co-autoria com Kasper Fisker e Peter Sepstrup, com a análise de uma ferramenta de apoio lexical para a escrita em língua estrangeira, um dos maiores desafios para a Lexicografia de todos os tempos, mas que ganha destaque com o avanço das tecnologias da informação. Como suprir as necessidades de usuários, que tradicionalmente combinariam consultas a vários tipos de dicionários (bilíngues e monolíngues) para produzir um texto, de modo mais personalizado (que responda às necessidades de expressão do consulente) e eficiente (sem que o consulente perca o foco de sua produção)? O artigo, densamente ilustrado, analisa a experiência de uso, por parte de aprendizes e professores de Inglês de diferentes contextos, do *Write Assistant*, ferramenta desenvolvida pela Ordbogen A/S (empresa dinamarquesa de TI), além de discutir a filosofia por trás da ferramenta, as técnicas aplicadas, sua base empírica e sua funcionalidade.

Também preocupada com a produção escrita, Ana Frankenberg-Garcia aprofunda o estudo sobre os tipos de exemplos em dicionários para aprendizes de Inglês como língua estrangeira. Partindo de seus trabalhos anteriores, distingue os exemplos entre aqueles que

facilitam a compreensão do sentido da palavra (a decodificação), e aqueles que fornecem elementos para a expressão oral e/ou escrita (a codificação), sobre os quais se concentra no artigo aqui apresentado. Com base em uma pesquisa experimental refinada envolvendo 161 graduandos portugueses de cursos diversos (Turismo, Hotelaria, Administração, Secretariado etc.), a autora comprova sua hipótese de que exemplos organizados com base em diferentes padrões léxico-gramaticais de uso facilitam a codificação em língua estrangeira. Resta saber qual seria a quantidade ideal de exemplos em um dicionário, aspecto mais dependente, conforme suas pesquisas, de preferências individuais dos aprendizes.

Trazemos também dois trabalhos que descrevem a elaboração de dicionários voltados mais especificamente à etapa escolar. O artigo de Rosa Estopà apresenta a construção de um dicionário escolar de ciências em língua catalã para crianças em fase de alfabetização, situando-se na fronteira dos estudos lexicográficos e terminológicos ou, mais precisamente, no âmbito da Terminologia Escolar. Partindo do pressuposto de que, para compreender o significado do que os professores falam, o que dizem os textos, testes e exercícios sobre ciências, as crianças devem poder se apoiar em suas próprias palavras, através de seus esquemas cognitivos, a autora desenvolve uma metodologia colaborativa e acumulativa. Entre outros procedimentos, constrói-se um *corpus* de representação do conhecimento das crianças a partir do que expressam, por meio de definições e desenhos, sobre determinado conceito científico, base para a elaboração do dicionário. O artigo comenta ainda um experimento posterior (uso do dicionário em um teste de compreensão) com 160 estudantes, mostrando como o trabalho com a linguagem pode andar junto com a construção do conhecimento científico. Já o artigo de Michèle Fourment-Berni Canani descreve em detalhes estratégias que ela criou para tornar a obra de sua autoria – um dicionário de aprendizagem bilíngue (francês > italiano) para adolescentes itálofonos – mais pedagógica do ponto de vista cultural. A abordagem intercultural atravessa todos os elementos do dicionário, fazendo do verbete o lugar privilegiado para tratar tanto das especificidades da cultura-alvo quanto da cultura compartilhada com a comunidade dos consulentes. Além disso, o texto ilustra, da mesma forma que seu dicionário, como descrição linguística e cultural são inseparáveis.

Lucía Marconi, Leonel Ruiz Miyares e Paola Cutugno apresentam uma pesquisa lexical do tipo básica, servindo tanto para aplicações pedagógicas quanto para trabalhos voltados à reabilitação da linguagem (com indivíduos disgráficos e/ou afásicos). Buscando caracterizar o léxico utilizado por estudantes cubanos – tanto em relação às formas lematizadas quanto às flexões –, o estudo analisa e compara dados extraídos do *Léxico Ativo-Funcional del Escolar Cubano* – feito com base em um *corpus* oral e escrito de mais de 700.000 palavras – e do *Diccionario Ortográfico del Español*. Com uma metodologia estatística sofisticada, o trabalho revela especificidades da variante do espanhol falado por jovens cubanos e estimula a realização de estudos descritivos em outros contextos.

Para fechar este número, contamos ainda com a tradução de dois artigos sobre a linguagem de gêneros científicos oriundos dos trabalhos desenvolvidos pelo projeto

Scientext (França). A partir da exploração de um *corpus* de língua francesa de textos científicos de diversas áreas do conhecimento, Francis Grossmann descreve o funcionamento de uma classe lexical intimamente relacionada à postura epistêmica da construção do conhecimento científico: os verbos “constativos”. Acorados no semantismo da percepção – diferentemente dos verbos de julgamento – verbos como *ver*, *observar*, *constatar*, *notar*, *perceber*, entre outros, adquirem um papel importante na escrita científica, mas guardam especificidades de uso condicionadas por seu perfil sintático e semântico, o que é analisado em detalhes nesse artigo. Por fim, Agnès Tutin trata da fraseologia transdisciplinar do artigo científico, apresentando os fundamentos teóricos e metodológicos do estudo do que podemos considerar como uma “linguagem de gênero”, ou seja, comum ao gênero discursivo em análise. Marcada por sequências lexicais que recobrem funções referencial, discursiva e interpessoal, além de rotinas retóricas (enunciados típicos), a linguagem do artigo científico em francês é descrita de uma perspectiva funcional, gerando um material que serve de base tanto para um estudo epistemológico do texto científico quanto para aplicações didáticas no contexto do letramento acadêmico.

Às autoras, aos autores e respectivas editoras nosso agradecimento e reconhecimento pela contribuição científica que realizam, pelo “jogo” que ensinam a jogar... Lembrando, por outro lado, que boa parte dos leitores deste volume do Cadernos de Tradução estaria sem boas condições de jogo se não fosse o trabalho primoroso de um time de peso: nossa(o)s estudantes e colegas do Instituto de Letras, responsáveis pelas traduções, supervisões e revisões dos textos, bem como pela colaboração na edição deste número, realizada pela bacharelada Sofia Froehlich Kohl. A toda(o)s nossa imensa gratidão.

Bora aprender como ensinar a jogar com as palavras! Boa leitura!

Sandra Dias Loguercio

Como citar este texto (ABNT):

LOGUERCIO, S.D. Apresentação. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 6-9, 2018.

**MÉTODOS E TÉCNICAS PARA USAR A INTERNET DIRETAMENTE
COMO CORPUS: O CASO DOS DICIONÁRIOS ON-LINE DE
ESPAÑHOL VALLADOLID-UVa^{1,2}**

Sven Tarp³

Pedro A. Fuertes-Oliveira⁴

Tradução: Luísa Rabaldo⁵

Revisão: Ana Eliza Pereira Bocorny⁶; Sandra Loguercio⁷

Resumo: Inicialmente, este artigo aborda algumas das consequências que o desenvolvimento tecnológico trouxe para a lexicografia, principalmente quanto aos diferentes tipos de bases empíricas que podem ser usados em projetos de dicionários. Em seguida, as principais vantagens e desvantagens de usar a Internet como corpus são listadas e comparadas com os resultados obtidos com corpora "tradicionais". Para ilustrar, o artigo mostra de que forma a Internet é usada como a principal fonte empírica para selecionar lemas e itens de significado para os Dicionários On-line de Espanhol Valladolid-UVa. Os métodos e ferramentas empregados no projeto são discutidos juntamente com as competências, conhecimentos e habilidades dos lexicógrafos. Por fim, o artigo fornece algumas conclusões gerais, bem como faz recomendações e levanta hipóteses para futuros trabalhos e pesquisas na área da lexicografia.

Palavras-chave: lexicografia da internet; lexicografia *online*; metodologia lexicográfica; base empírica; seleção do lema; seleção do significado; bases de dados lexicográficos; dicionários de espanhol; dicionários de língua geral.

1. Introdução

Por meio da análise do processo geral de compilação de dicionários descrito por Fuertes-Oliveira e Tarp (2014: 85), conclui-se que há três casos em que os lexicógrafos

¹ N.A.: Este artigo é uma adaptação de um texto publicado em *Lexikos* 26 (TARP; FUERTES-OLIVEIRA, 2016). N.E.: A presente tradução foi autorizada para ser publicada em português pelo autor Sven Tarp.

² Tradução revisada pelo autor Sven Tarp e título modificado conforme solicitação do autor.

³ Centro de Lexicografia, Universidade de Aarhus, Dinamarca e Departamento de Africâner e Dinamarquês, Universidade de Stellenbosch, África do Sul (st@cc.au.dk).

⁴ Centro Internacional de Lexicografia, Universidade de Valladolid, Espanha e Departamento de Africâner e Dinamarquês, Universidade de Stellenbosch, África do Sul (pedro@emp.uva.es).

⁵ Bacharelada em Letras Português/Inglês e bolsista de Iniciação Científica, Instituto de Letras, UFRGS.

⁶ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁷ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

podem necessitar de acesso a dados empíricos para fazer um bom trabalho. O primeiro é quando procuram informações sobre as necessidades lexicográficas do público-alvo, com o objetivo de definir uma concepção de dicionário que possa responder às suas necessidades; o segundo é quando selecionam e preparam os dados lexicográficos para inclusão em um dicionário; e o terceiro é quando avaliam a utilidade do dicionário em termos de satisfação do usuário. Na verdade, há também uma quarta situação em que dados empíricos externos podem ser necessários: quando lexicógrafos avaliam o mercado para determinar as possibilidades de vendas do produto. O quarto caso, entretanto, está mais relacionado ao lado comercial do projeto do que aos aspectos lexicográficos, em um sentido estrito. De qualquer forma, para cada uma dessas situações existe um conjunto de métodos mais ou menos apropriados, isto é, mais ou menos confiáveis e rápidos em termos de produtividade e qualidade do produto final.

Em seguida, veremos as bases empíricas e os métodos correspondentes que podem ser aplicados ao selecionar lemas e itens de significado (sentido) em um projeto lexicográfico *on-line*. Discutiremos algumas das vantagens e desvantagens mais importantes ao usar a Internet como um *corpus* e compararemos com os resultados obtidos com *corpora* “tradicionais” de texto. Como exemplo, utilizaremos um projeto que está sendo realizado atualmente no Centro Internacional de Lexicografía da Universidade de Valladolid, chamado *Diccionarios en Línea de Español “Universidad de Valladolid”*, referidos neste artigo como Dicionários On-line de Espanhol Valladolid-UVa. O projeto, que originalmente foi iniciado como uma colaboração entre o Centro de Valladolid e o Centro de Lexicografía da Universidade de Aarhus, baseia-se na teoria funcional da lexicografía e é inspirado em um projeto dinamarquês similar (ver Fuertes-Olivera; Bergenholtz, 2015). Por fim, apresentaremos os resultados obtidos até o presente momento e abordaremos a necessidade de contar com a intuição como um método intangível, mas muito relevante e inevitável no desenvolvimento de dicionários.

2. A relação entre lexicografía e tecnologia

Em uma perspectiva histórica, pode-se observar uma relação íntima e complexa entre a lexicografía e a tecnologia. Isso implica, entre outras coisas, que o desenvolvimento tecnológico pode não apenas criar novas ferramentas para auxiliar lexicógrafos na realização de tarefas, mas também novas bases empíricas a partir das quais podem recuperar dados, bem como depararem-se com a necessidade e a possibilidade de desenvolverem novos métodos de trabalho. A reflexão é particularmente relevante em períodos históricos como o atual, em que novas tecnologias disruptivas estão sendo introduzidas na lexicografía com consequências que ainda não são totalmente compreendidas:

Atualmente estamos no meio de uma nova transição da base material e tecnológica da lexicografía com a introdução de novas ferramentas e métodos de produção, bem como novas plataformas

e meios para apresentar o produto lexicográfico e o uso extensivo de *corpora* para a coleta de material. O desenvolvimento e a inovação tecnológica estão mais acelerados do que nunca. (...) Sabemos o ponto de partida, mas temos apenas uma vaga ideia de onde chegaremos. (Gouws; Tarp, 2017)

Em geral, há uma variedade de fontes das quais os lexicógrafos podem obter dados. Bergenholtz e Tarp (1995: 90-96) discutem, dentre as mais importantes, a introspecção, a multispecção, os especialistas externos, os dicionários existentes, os manuais, os livros didáticos, os cartões de exemplo e os *corpora* textuais. Com exceção das três primeiras, essas fontes empíricas só são possíveis graças ao desenvolvimento tecnológico em vários estágios: invenção do papel, canetas, encadernação de livros, máquinas de impressão, computadores e bancos de dados. Desde então, com a introdução e o desenvolvimento da tecnologia *on-line*, outra fonte de dados empíricos foi colocada à disposição da lexicografia: a Internet.

É interessante notar que Bergenholtz e Tarp (1995) discutem a consulta a especialistas externos como uma forma de multispecção e o uso do próprio conhecimento como uma forma de introspecção. Pode ser que tenham razão, mas parece que há uma diferença entre o uso da introspecção em termos de habilidades e competências linguísticas, como é normalmente compreendido dentro da linguística, e o uso de conhecimento especializado e atualizado que está armazenado na memória de alguém. Como Tarp (2008: 131-136) argumentou, no desenvolvimento de vários tipos de dicionários, é importante distinguir, de um lado, as habilidades linguísticas e, de outro, o conhecimento erudito de uma língua, tal como fornece, por exemplo, a teoria linguística. A esse respeito, parece lógico também distinguir o uso da competência de linguagem e o uso de conhecimento especializado no processo de compilação de dicionários. Assim, embora exista certa confusão terminológica na literatura lexicográfica existente, a introspecção - em vez de ser uma base empírica em si - deve ser considerada um método para a obtenção de tipos específicos de dados empíricos. Trata-se de um método feito para "olhar" para si mesmo, a fim de adquirir material para diferentes propósitos. As bases empíricas "internas" das quais os lexicógrafos podem se valer por meio desse método são as competências, habilidades e conhecimentos de linguagem, aos quais pode-se acrescentar a experiência pessoal em geral.

As diversas fontes empíricas raramente são usadas sozinhas. Em uma resenha de livro, Kilgariff (2012) fornece um exemplo sobre como dois tipos diferentes de bases empíricas combinam-se:

Eu notei um erro lexicográfico. Nas páginas 211-213, temos uma análise do verbo frasal em inglês *call back*. Seis significados são apresentados e, no sexto, surge o exemplo "*I cannot call his face back.*" Como falante nativo de inglês, esse exemplo me causou

estranheza. Essa frase está completamente errada. (Poderíamos dizer *I cannot recall his face*). Uma breve pesquisa revelou que a tal "frase exemplo" existe em vários dicionários e ferramentas de tradução: um erro dicionarizado que foi copiado e recopiado de dicionário para dicionário. (Kilgarrieff 2012: 28)

Ao dizer que “esse exemplo me causa estranheza” como falante nativo de inglês, Kilgarrieff sugere que sua competência na língua materna o alerta sobre um possível problema, confirmado e explicado por meio da consulta de outras bases empíricas, neste caso, dicionários existentes e ferramentas de tradução. Esse é claramente o método correto a ser aplicado em tais situações, uma vez que “a fonte primária de evidências do lexicógrafo sobre como uma palavra é usada muda de subjetiva para objetiva; da introspecção à observação dos contextos” (Kilgarrieff 1997: 111). Nota-se que Kilgarrieff fala aqui da fonte primária de evidência, que não é a única, embora erre ao definir introspecção e observação dos contextos como fontes, na medida em que ambos são métodos para acessar as fontes reais de evidência.

3. Corpus vs. Internet: discussão preliminar

Os primeiros *corpora* compostos com textos eletrônicos foram introduzidos na década de 1960, e não pararam de crescer desde então. As duas primeiras décadas após o surgimento foram caracterizadas por uma significativa batalha de ideologias entre os pesquisadores que defenderam a relevância dos *corpora* tanto para a linguística como para a lexicografia, e aqueles que se opuseram a essa visão com vários argumentos, geralmente defendendo a introspecção como método mais apropriado para a obtenção de material empírico. Um dos defensores da introspecção foi Lees (1962) que declarou:

Você é um falante nativo de inglês; em dez minutos é capaz de produzir mais ilustrações de qualquer questão gramatical da língua inglesa do que achará em milhões de palavras em textos aleatórios. (Lees 1962: 110)

Aos poucos, a discussão desapareceu. Meio século após sua introdução, não resta mais nenhuma dúvida de que *corpora* de textos eletrônicos podem ter grande valor não apenas para a pesquisa linguística, mas também para os lexicógrafos ao realizarem uma série de tarefas relacionadas à compilação de dicionários. Isso tem sido debatido por vários estudiosos envolvidos na lexicografia prática, entre eles Bergenholtz (1996), Atkins e Rundell (2008) e Hanks (2012). A prova do pudim é a existência de muitos dicionários de alta qualidade que foram compilados fazendo uso desse tipo de base empírica (ver, por exemplo, Sinclair, 1997), embora a ânsia tenha ido, por vezes, longe demais, pelo menos quanto à seleção de termos e definições em dicionários especializados (ver Tarp, 2016, e Xue; Tarp, 2016).

No entanto, como consequência negativa desse desenvolvimento positivo, a introspecção como método de uso das competências e conhecimentos próprias de um indivíduo é, por vezes, subestimada ou até mesmo ignorada. Embora “lexicógrafos nunca devam confiar somente na abordagem introspectiva” (Bergenholtz; Tarp 1995: 92), especialmente em casos dúbios, frequentemente esquece-se que a introspecção sempre existe como um filtro na base das escolhas do lexicógrafo, na medida em que nenhum dicionarista introduziria dados linguísticos ou de qualquer outro tipo com os quais discorda – como a estranheza causada em Kilgarriff – sem antes negociar sua correção com outras fontes empíricas.

Hoje em dia, *corpora* compostos de textos contendo centenas de milhões de palavras estão disponíveis aos compiladores de dicionários. O *Big Data*, por exemplo, já é uma realidade, mas o entusiasmo gerado por esse desenvolvimento não deve ofuscar o fato de que nenhum *corpus*, por maior que seja, pode competir com a enorme coleção de textos e palavras que podem ser acessados através da Internet. O desenvolvimento de métodos que permitam o uso dessa base empírica quase ilimitada constitui, sem dúvida, um desafio cada vez mais relevante para a lexicografia.

Segundo Fuertes-Olivera (2014: 51), um *corpus* lexicográfico, ou seja, um *corpus* que pode ser usado para auxiliar na criação de dicionários, pode ser definido como “qualquer coleção de textos em que os lexicógrafos possam encontrar inspiração para completar as estruturas das quais precisam na construção de um dicionário real”. Como já mencionado, a Internet é composta por uma coleção de textos. Assim, se um lexicógrafo pode encontrar inspiração nessa grande coleção de textos, a Internet também pode ser considerada um tipo de *corpus* lexicográfico de acordo com a definição acima. Esse é também o ponto de vista de Kilgarriff e Grefenstette (2003: 334) que dizem que “a resposta à questão 'a Internet é um corpus?' é sim.”

Assim, há duas maneiras diferentes de usar a Internet em um projeto lexicográfico, a saber: 1) construir um *corpus* de textos encontrados na Internet, e 2) usar a Internet diretamente como um *corpus* - em ambos os casos por meio de ferramentas de busca e outros recursos. Cada um desses dois tipos de *corpus* lexicográfico tem suas vantagens e desvantagens. Abaixo listamos algumas das vantagens do uso da Internet como *corpus*, em relação ao uso de *corpora* “tradicional” composto de coleções de textos, sejam coletados da Internet ou de outro lugar:

- Os lexicógrafos têm acesso a um número muito maior de textos do que os incluídos em qualquer *corpus* de textos selecionados.
- Os textos estão sempre atualizados.
- Tempo e dinheiro são poupados quando não é necessário compor um *corpus* (que é um requisito em relação a tipos específicos de dicionários, principalmente os especializados).
- O processo de busca pode ser facilmente limitado a áreas geográficas específicas, um fato importante para um idioma multinacional como o espanhol.

- O uso da Internet pode levar à identificação e à seleção de mais itens de significado do que aquelas que podem ser encontradas em um *corpus* fabricado.

Quanto às desvantagens de usar a Internet como *corpus*, os seguintes aspectos parecem ser os mais importantes:

- Nem a qualidade nem a origem dos textos podem ser controladas.
- Em alguns textos, os autores podem não ser pessoas reais.
- Os autores podem ter um baixo nível de proficiência na língua em questão.
- Os textos podem não ter sido revisados e corrigidos.
- É difícil calcular a frequência dos fenômenos linguísticos que aparecem nos textos.

Algumas das desvantagens citadas acima podem não ser relevantes para projetos concretos de dicionários. Gudmann (2014: 32), por exemplo, argumenta que “as informações sobre frequência (...) não são particularmente relevantes para um dicionário geral monolíngue”. Em outros casos, as desvantagens podem ser neutralizadas ou, ao menos, consideravelmente reduzidas, por um lexicógrafo bem preparado que desempenhe um papel ativo com base em sua competência linguística, suas habilidades, seu conhecimento e sua experiência. Retomaremos essa questão mais adiante. Até aqui, nossa conclusão preliminar é que, apesar das inegáveis desvantagens, é perfeitamente possível, e até mesmo benéfico, usar a Internet como principal fonte empírica, sem recorrer aos *corpora* “tradicionais” de textos, quando o objetivo é a produção de dicionários com qualidade ainda maior.

4. Selecionando lemas

Nós, sinceramente, duvidamos que o *corpus* tradicional, composto por uma coleção de textos, seja a fonte empírica mais apropriada para a seleção de lemas, principalmente se essa seleção for feita desde o início. Até onde sabemos, os grandes dicionários de língua geral que usam *corpora* para esse propósito são, em sua maioria, dicionários que tinham um estoque de lemas selecionado antes da introdução dos *corpora*, que agora são usados “apenas” para fornecer lemas adicionais, entre outros dados. Um método diferente e uma base empírica são, portanto, necessários quando o desafio é uma seleção rápida e confiável de lemas para um projeto lexicográfico completamente novo da magnitude dos Dicionários On-line de Espanhol Valladolid-UVa, planejados para lidar com mais de cem mil lemas e um número muito maior de significados. A principal base empírica escolhida para esse projeto foi, portanto, a Internet, acompanhada de um método que será descrito nesta seção.

A ideia básica é que a Internet já contém um número considerável de listas de palavras menores ou maiores, com acesso e uso gratuitos. O desafio é, portanto, encontrar essas listas e fazer uso delas. Isso é feito por meio de um rastreador de Internet projetado especificamente para esse fim pela empresa dinamarquesa Ordbogen.com, que, devido ao seu modelo de negócios, é o provedor de dicionários *on-line* em assinatura mais bem-sucedido do mundo atualmente.

Assim que determinadas listas de palavras úteis são encontradas pelo rastreador de Internet, essas listas são copiadas e coladas em um chamado carregador de lema (veja a Figura 1), outra ferramenta desenvolvida pela Ordbogen.com e concebida pelo professor Henning Bergenholtz do Centro de Lexicografia de Aarhus. O carregador de lema atribui automaticamente um lema a um cartão no banco de dados e tem a vantagem de não duplicar a entrada, rejeitando-os caso já constem no banco de dados.

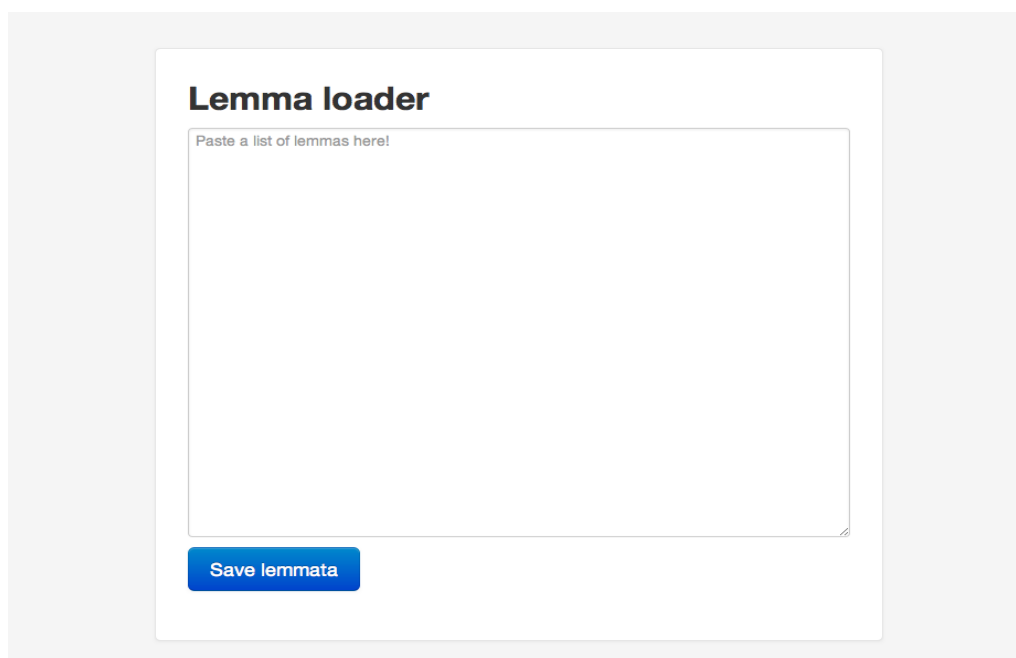


Figura 1: Captura de tela do carregador de lema com o campo em que as listas de palavras copiadas são coladas.

As experiências mostram que o método para selecionar lemas por meio de um rastreador de Internet e um carregador de lema é muito eficiente, rápido e totalmente confiável no caso de lemas compostos por uma palavra. Isso é revelado pelo fato de que, apenas um mês após o início da seleção de lemas para os Dicionários On-line de Espanhol Valladolid-Uva (julho de 2013), o banco de dados já continha 58.000 cartões com lemas de uma palavra.

O passo seguinte no processo de seleção de lema é a revisão manual que ocorre quando os dados gramaticais formais são anexados ao cartão do lema. A revisão é realizada pelo editor-chefe, Pedro A. Fuertes-Olivera. Devido às características dos dicionários e à capacidade de armazenamento quase ilimitada do banco de dados, o projeto não trabalha com critérios de inclusão de lema, apenas com critérios de exclusão. Com essa abordagem, apenas os lemas que apresentam claramente erros ortográficos ou que não podem ser documentados na base empírica, isto é, na Internet, são excluídos, mesmo quando são encontrados em algumas listas de palavras antigas, mas não na Internet como tal. Até agora foram poucos os casos em que um lema teve de ser excluído, o que também aponta para a eficiência do método.

A fim de garantir um tratamento sistemático da língua, após a seleção inicial foram elaboradas várias listas temáticas com cores, números, cidades de determinado tamanho, rios de mais de 1.000 km, etc., e as palavras correspondentes introduzidas no banco de dados como lemas. Esse trabalho durou de setembro a novembro de 2013 e resultou na inclusão de mais 10.000 lemas.

Além das fontes empíricas mencionadas, outras fontes também são usadas para fornecer um fluxo de novos lemas. Por exemplo, quando os lexicógrafos estão consultando a Internet para identificar itens de significado para os lemas selecionados (ver Seção 5), eles detectam simultaneamente um número considerável de sinônimos, antônimos e combinações de palavras que são continuamente introduzidos no banco de dados como novos lemas.

Por fim, há a questão de expressões idiomáticas e outras expressões fixas que são geralmente selecionadas como lemas independentes nos Dicionários On-line de Espanhol Valladolid-UVa. Neste caso, há quatro fontes: 1) No processo de detecção de itens de significado, ocasionalmente aparecem expressões fixas. Nesse caso, elas são enviadas ao editor-chefe, que as analisa e avalia pesquisando na Internet. 2) Dicionários existentes também são usados como fontes, 3) assim como o CREA Corpus composto e publicado pela *Real Academia Española*, de uso livre. 4) Por fim, várias expressões fixas também são encontradas em outras fontes, por exemplo, livros e artigos lidos pelos lexicógrafos relacionados a outras tarefas.

Como se pode ver, apenas as últimas três fontes de busca de expressões fixas não se apoiam na Internet como base empírica para a seleção de lemas para os Dicionários On-line de Espanhol Valladolid-UVa. Geralmente, o processo é muito rápido, eficiente e de baixo custo, o que, até agora, resultou em cerca de 20% mais lemas do que aqueles contidos nos dicionários espanhóis, até então, maiores.

5. Selecionando itens de significado

O método desenvolvido para selecionar itens de significado nos Dicionários On-line de Espanhol Valladolid-UVa é fortemente inspirado em um método similar usado no *Danish Internet Dictionaries* (ver Bergenholtz; Agerbo, 2014), mas tem algumas particularidades próprias. Grosso modo, o método de seleção de significado engloba as seguintes quinze etapas:

1. Um lema contido no banco de dados é escolhido na interface do usuário do lexicógrafo (ver Figura 2);
2. O botão “Google”, localizado à esquerda na interface do lexicógrafo, é ativado;
3. Um resultado “tradicional” da pesquisa do Google é exibido (ver Figura 3);
4. As primeiras (3-20) páginas são ignoradas porque contêm apenas dados irrelevantes, em termos lexicográficos;
5. Os mini-textos exibidos em cada página são lidos para que se obtenha uma ideia geral do que se trata;

6. Usando o método “copiar e colar”, as partes relevantes dos mini-textos são copiadas para um documento do Word;
7. Simultaneamente, colocações, exemplos, sinônimos, antônimos e combinações de palavras são selecionados para serem introduzidos nos respectivos campos no cartão representando o sentido em questão na interface do lexicógrafo (ver Figuras 4 e 5). Expressões idiomáticas e expressões fixas são enviadas ao editor-chefe para avaliação adicional;
8. Várias páginas do Google são revisadas até que não se encontre mais dados novos e tudo seja repetido. O número de páginas depende das características de cada lema, assim como da intuição do lexicógrafo, que se baseia na sua experiência;
9. Assim que uma quantidade satisfatória de dados empíricos é selecionada, os dados são agrupados de acordo com o significado;
10. Com base nos grupos de dados, as primeiras definições são escritas de acordo com as instruções lexicográficas preparadas pelo editor-chefe;
11. Nesta etapa, o lexicógrafo decide se está satisfeito, ou se é necessário repetir o processo inteiro ou parte dele para obter uma quantidade satisfatória de evidências empíricas;
12. Assim que o lexicógrafo termina a seleção de significados e redige as definições relativas a um lema, uma mensagem é enviada ao editor-chefe;
13. O editor-chefe revisa as definições e as compara com as que aparecem em quatro dicionários de espanhol previamente selecionados (ver Seção 6). Se alguma informação estiver faltando, pode ser que se inicie um novo processo de pesquisa, pois um princípio básico do projeto é que nenhuma definição seja copiada de outros dicionários;
14. Se as definições estiverem relacionadas a termos que ocorrem também na linguagem cotidiana, especialistas externos podem ser consultados para definir com mais exatidão;
15. Quando o editor-chefe estiver satisfeito - e outros dados relevantes, como gramática, sinônimos, antônimos, formações de palavras, colocações e exemplos forem incluídos - o lema é indicado para publicação *on-line*.

Até julho de 2016, o banco de dados dos Dicionários On-line de Espanhol Valladolid-UVa contém cerca de 40.000 cartões finalizados (cada um deles representando um significado), prontos para publicação. A experiência até agora mostra que, para 70% dos lemas, trabalhar com os mini-textos que aparecem como resultado da pesquisa do Google é suficiente. Para os 30% restantes, é necessário, ainda, acessar um ou mais *links* para encontrar dados adicionais em outros documentos. Neste último caso, os dados necessários são encontrados em 90% das situações. Apenas em 10% das ocorrências, o que representa 3% da totalidade dos lemas, é necessário realizar uma nova busca com uma variante do lema procurado. Isso significa que, para 97% de todos os lemas, uma única pesquisa do Google é suficiente para obter o material empírico necessário para a seleção de itens de significado e redigir definições de acordo com os critérios estabelecidos.

25.154 / 121.524 →
25154

Clase de palabra
Sinónimo
Colocación
Polisemia
Encontrar
Google
Diccionario
Nuevo
Copiar
Borrar
Guardar

pedro finito
eva finito

cachupina Informal Hom Pol

Neutro

Borrar Primera ref. Borrar Ref. gr.

Nota gramatical

Borrar nombre <una cachupina, la cachupina, unas cachupinas, las cachupinas> Prioridad: 1

una cachupina la cachupina unas cachupinas las cachupinas Gram:

nombre	nombre propio	verbo	adjetivo	pronombre	artículo	expresión
conjunción	adverbio	interjección	numeral	abreviatura	preposición	dicho
un -, el -, unos -s, los -s	una -, la -, unas -s, las -s	presente	participio	-s		
un -, el -, unos -es, los -es	una -, la -, unas -es, las -es	pasado	gerundio	-es		
	un -, el -, unas -s, las -s	imperfecto	sin participio			
sin flexión	sin singular, unas -, las -	futuro				
sin singular, unos -s, los -s		condicional				
sin singular, unos -es, los -es		presente subjuntivo				

Figura 2: Interface do lexicógrafo da gramática da palavra “cachupina”.

Google cachupina

Alle Billeder Maps Videoer Mere ▼ Søgerværktøjer

Side 10 af ca. 8.350 resultater (0,31 sekunder)

CVC. El legado de Sefarad. Memoria de Sefarad en América. La niña ... ✓
cvc.cervantes.es > ... > Memoria de Sefarad en América ▼ Oversæt denne side ✓
Otras veces se la llamaba «cachupina» o «refugacha», como si ella hubiera cometido los pecados de los conquistadores de la época colonial. Se acostumbró a ...

Las gafas del enemigo -Hispano Post ✓
www.hispanopost.com/las-gafas-del-enemigo ▼ Oversæt denne side ✓
23. maj 2016 - Libertad y garantías para la Cachupina de @saludhernandezm Ya. #SaludHernandez #DondeEstaSaludHernandez. Please enable ...

en el cerro de la culebra, el cíclope es ecocida - EnlaceVeracruz 212 ?
archivo.vazquezchagoya.com/?p=4748 ▼ Oversæt denne side ✓
24. jul. 2008 - La Cachupina, tenía un dolor intenso y sentí angustia porque no podía caminar, ni siquiera dar un paso. Mi laptop estaba ausente y pedí favor ...

Los Andes-Pelluhue Chile Slideshow & Video | TripAdvisor™ ✓
tripwow.tripadvisor.com/tripwow/ta-0743-9a6a-58ec ▼ Oversæt denne side ✓
Matty- Katty- Patricia- Cachupina. Matty- Katty- Patricia- Cachupina - Mis hijos y los hijos de mis hijos! Mis hijos y los hijos de mis hijos! Skiing - Portillo. Skiing - ...

Foto de COROZAL (SUCRE), Colombia ✓
www.fotopaises.com/foto/85027 ▼ Oversæt denne side ✓
... publicar unas... gracias. Anónimo, Publicado: 24/3/2015 a las 4:32 por Usuario anónimoEn nuestro vocabulario popular le decimos, la flor de la cachupina.

Figura 3: Resultado da pesquisa no Google pela palavra “cachupina”.



Figura 4: Interface do lexicógrafo para a introdução de definição, sinônimos e antônimos para a palavra “cachupina”.

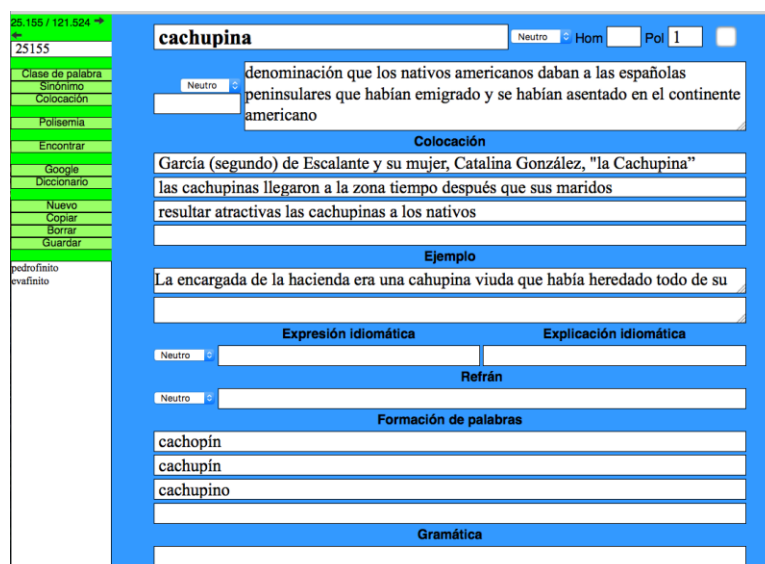


Figura 5: Interface do lexicógrafo para introdução de colocações, frases de exemplo, expressões fixas e formações de palavras para a palavra “cachupina”.

6. Comparação com dicionários de espanhol semelhantes

Como mencionado acima, depois de redigir as definições dos diferentes significados, eles são comparados com aqueles encontrados em quatro dicionários de espanhol, a saber:

- María Moliner: Diccionario de Uso del Español (DUE)
- Aquilino Sánchez Pérez: Gran Diccionario de Uso del Español Actual (GDUEA)
- Manuel Seco: Diccionario del Español Actual (DEA)

- Real Academia Española: Diccionario de la Lengua Española (DLE)

Estes quatro dicionários estão entre os maiores e mais prestigiados dicionários de língua geral do espanhol. A comparação também serve como um tipo de controle de qualidade e indicação do que poderia ser melhorado (ver passo 13 na Seção 5). Até agora, a comparação tem sido favorável aos Dicionários On-line de Espanhol Valladolid-UVa, pois mostra que cada lema tratado com o método descrito tem uma média de 30-40% mais significados do que os lemas encontrados nos outros quatro dicionários.

A Tabela 1 mostra o número de significados que os cinco dicionários mencionados fornecem para oito lemas diferentes. Esses dados não são de forma alguma representativos, mas apenas um indicativo de como o método descrito na seção anterior, em alguns casos, pode gerar um número maior de significados.

Tabela 1: Comparação entre cinco dicionários de espanhol quanto ao número de significados de oito lemas selecionados.

Palavra	DUE	GDUEA	DEA	DLE	Valladolid
ababol	1	1	1	2	3
Cabila	1	1	2	2	3
Cable	4	4	4	6	11
cabestro	4	4	2	4	6
eclipsar	2	2	2	2	3
eclipsarse	2	2	2	3	4
halagar	4	2	2	4	3
machaca	5	5	5	6	15

A tendência refletida na Tabela 1 é corroborada por Gudmann (2015), que estudou cinco dicionários *on-line* de espanhol - entre eles a obra da Real Academia Española - e identificou um número surpreendentemente grande de lacunas de significado. Outra ilustração deste fenômeno é o tratamento das palavras *cachupín* e *cachupina*, que são usadas em algumas partes da América Latina e apresentadas como um único lema nos quatro dicionários mencionados acima, cada um deles com apenas um sentido. A Figura 7 mostra como eles são abordados pela Real Academia Española na versão *on-line* do Diccionario de la Lengua Española:

cachupín, na

Del dim. del port. *cachopo* 'niño'.

1. m. y f. despect. **gachupín**.

Real Academia Española © Todos los derechos reservados

Figura 6: O lema “cachupín, na” no Diccionario de la Lengua Española.

Este modo de apresentar as duas palavras pode ser, com razão, considerado sexista, como se a palavra feminina *cachupina* fosse apenas subordinada à palavra masculina *cachupín*. Nos Dicionários On-line de Espanhol Valladolid-UVa, as duas palavras são tratadas separadamente e listadas como dois lemas diferentes. Isso também sugere que cada uma das duas palavras foi pesquisada separadamente no Google, que busca itens de significado. O resultado surpreendente desse método é que a palavra masculina aparece com dois sentidos, enquanto a palavra feminina inclui nada menos que oito sentidos, ou seja, um total de dez sentidos, como pode ser visto nas capturas de tela do dicionário extraído do banco de dados e apresentadas nas Figuras 7 e 8.

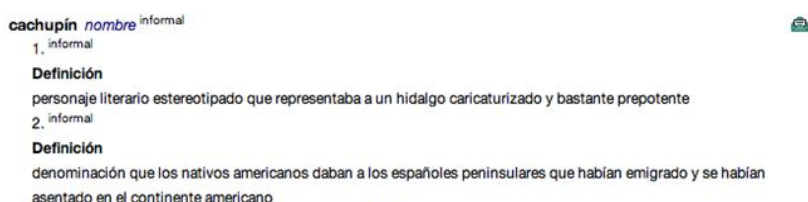


Fig. 7: O lema “cachupín” nos Dicionários On-line de Espanhol Valladolid-UVa.

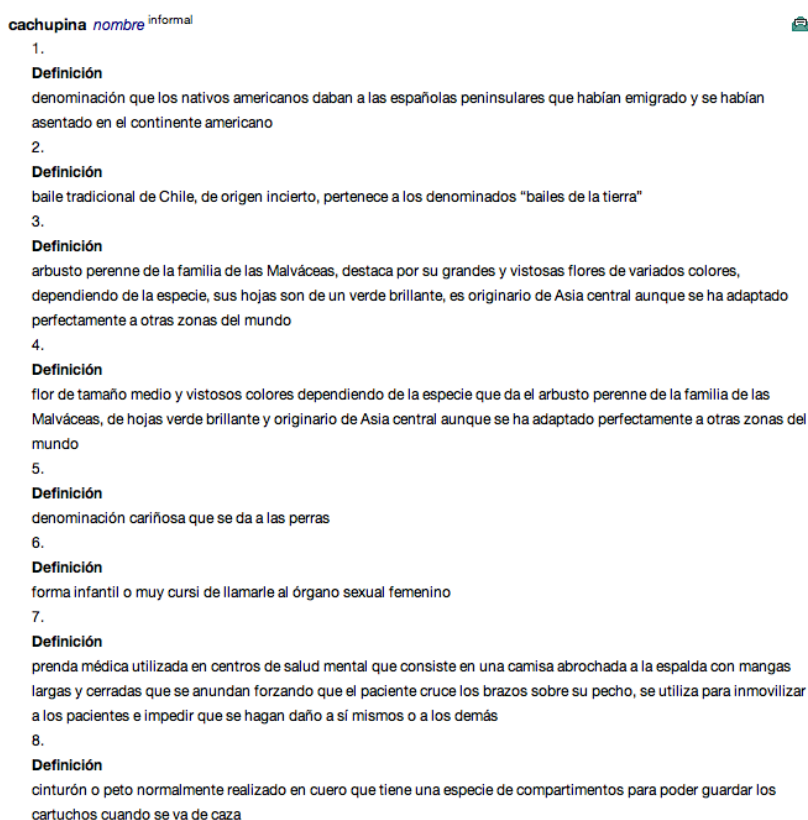


Figura 8: O lema “cachupina” nos Dicionários On-line de Espanhol Valladolid-UVa.

As diferenças acima não devem ser absolutizadas, pois o número de lemas e sentidos coletados no banco de dados dos Dicionários On-line de Espanhol Valladolid-UVa não pode ser comparado diretamente com os outros quatro dicionários do espanhol.

Há duas razões para isso. A primeira é que os quatro dicionários analisados são impressos, e, por isso, sofrem com as restrições de espaço. A segunda é que a “filosofia do papel” continua a influenciar a atual lexicografia espanhola, mesmo no ambiente virtual. Em relação a isso, um problema sério é que os critérios de seleção ainda não foram adaptados à nova tecnologia. Rundell (2015) mostra, de forma condensada, como esses critérios mudaram frente a desenvolvimentos recentes:

Então, quando não há restrições de espaço, pode ser que faça mais sentido mudar a pergunta e, em vez de questionar, "esta palavra passa nos meus testes de inclusão?", devemos perguntar, "há razões plausíveis para não incluir essa palavra?" (Rundell 2015: 312)

Esta mudança nos critérios de seleção sugere que os lexicógrafos, que, há apenas algumas décadas, deviam justificar a inclusão de qualquer novo lema ou sentido - porque, normalmente, significava a exclusão de outros dados -, são agora desafiados com a necessidade de justificar a não-inclusão de novos lemas e acepções. A esse respeito, não sabemos quantas acepções ainda não publicadas os outros dicionários espanhóis podem ter em seus bancos de dados. Também não sabemos quantos itens de significado adicionais poderiam ser identificados em seus *corpora*, se fosse necessário. A comparação entre os cinco dicionários apresentados na Seção 6 deve ser compreendida como um indicativo das novas possibilidades que o uso da Internet coloca à disposição da lexicografia, e não como um fato baseado em evidências que explica todos os aspectos. No final do dia, o que deve ser comparado não são os dicionários em si, mas as bases e os métodos empíricos usados para fornecer os dados lexicográficos.

Com isso em mente, nossa hipótese atual, em fase de submissão para outras pesquisas, é que um *corpus* composto de textos selecionados é mais apropriado para identificar as palavras, os sentidos e os comportamentos das palavras mais comuns e mais frequentes, enquanto o uso da Internet diretamente como um *corpus* é um método mais apropriado quando há a necessidade de detectar palavras, expressões, sentidos e comportamentos menos típicos e frequentes das palavras.

7. Qualidade e produtividade

Como resultado da discussão anterior, conclui-se que o uso da Internet diretamente como um *corpus* representa um método promissor para explorar dois resultados relevantes provindos do desenvolvimento tecnológico, a saber: 1) que a Internet hoje compreende um número quase “ilimitado” de textos e palavras; 2) que um dicionário digital moderno é sustentado por um banco de dados com capacidade de armazenamento quase “ilimitada”. A escolha do método é, portanto, um tópico interessante para discussões acadêmicas e uma questão que tem consequências práticas e econômicas consideráveis. Atualmente, o desafio para as editoras e equipes lexicográficas não é apenas compilar dicionários de alta qualidade, mas também garantir alta produtividade no

processo de compilação. Cada vez mais, os lexicógrafos estão se dando conta de que esse ofício está imerso em uma crise que, de certa maneira, pode ser descrita como uma luta entre a vida e a morte. Essa crise é determinada por duas tendências opostas na lexicografia atual: por um lado, muitos editores de dicionários de alta qualidade estão fechando seus departamentos de dicionário devido à falta de renda e a um modelo de negócios sustentável. Por outro lado, um número crescente de dicionários de livre acesso, mas de qualidade duvidosa, é colocado na Internet por pessoas geralmente bem intencionadas mas sem treinamento suficiente.

Os resultados dessa evolução são muitos e, em sua maioria, negativos. Embora os dicionários, devido à Internet, tenham mais usuários do que nunca, muitos desses usuários enfrentam problemas quando tentam utilizar as informações retiradas de dicionários duvidosos, um fato que muitos professores de línguas e de tradução reconhecerão. O resultado inevitável é que um número cada vez maior de usuários, conscientes de suas necessidades, recusa a lexicografia e procura outras ferramentas para obter informações. Algumas grandes empresas espanholas que estão dispostas a pagar pelo serviço são, por exemplo, críticas ao padrão atual de dicionários *on-line* de espanhol, deixando um mercado aberto para projetos como o Valladolid-UVa. No entanto, o problema é ainda mais exacerbado pelo fato de que muitos usuários da Internet, especialmente os jovens, esperam que o serviço seja gratuito, como discutido por Gouws & Tarp (2017).

Editoras de todo o mundo estão lutando para encontrar uma solução para esses complexos desafios. Uma das contra-medidas necessárias para a crise atual é, sem dúvida, aumentar a produtividade no processo de compilação do dicionário, visando reduzir custos e encontrar um modelo de negócio sustentável – ou seja, aumentar a produtividade sem comprometer a qualidade. A produtividade tem muitas faces e só pode crescer por meio da integração de tecnologia amigável, de métodos eficientes e de lexicógrafos bem treinados e motivados. Em projetos de dicionários modernos como o Valladolid-UVa, a interface do lexicógrafo é a ferramenta central de trabalho por meio da qual eles introduzem os dados no banco de dados. A esse respeito, Tarp (2015) escreve:

A interface do lexicógrafo é basicamente um meio de produção. Por conseguinte, deve ser concebida visando garantir uma produtividade elevada e a mais alta qualidade possível do produto, ou seja, dos dados armazenados na base de dados. Para isso, é necessário, acima de tudo, que contenha todos os campos necessários para introduzir dados lexicográficos dos tipos previstos na base de dados. Mas também é importante que a interface seja tão amigável quanto for possível, a fim de facilitar o trabalho do lexicógrafo, reduzir o número de erros, economizar nos recursos empregados e encurtar o tempo total de produção. (Tarp 2015: 234)

As Figuras 2, 4 e 5 da Seção 6 mostram três capturas de tela da interface do lexicógrafo relacionada ao lema *cachupina*. O primeiro representa o cartão-mãe no qual os dados gramaticais comuns a todos os sentidos do lema são introduzidos. A segunda e terceira captura de tela mostram duas páginas do cartão representando o primeiro sentido do lema. As páginas são estruturadas de acordo com as diferentes tarefas que devem ser executadas e de tal forma que a necessidade de alternar entre uma e outra por meio dos botões funcionais à esquerda seja reduzida ao máximo. A ideia principal é que os lexicógrafos se sintam confortáveis ao trabalhar com essa interface. Seu *design* de fácil utilização, juntamente com os métodos descritos nas seções anteriores, é condição para a alta produtividade que caracteriza a compilação dos Dicionários On-line de Espanhol Valladolid-UVa.

A identificação de itens de significado e a introdução de definições e outros dados lexicográficos no banco de dados começaram em março de 2014. A experiência mostra que um lexicógrafo pode finalizar uma média de 4 a 6 acepções por hora com o método descrito e a ferramenta de produção. A experiência também mostra que a produtividade diminui após quatro ou cinco horas, pois o trabalho exige um alto grau de concentração – razão pela qual os quatro lexicógrafos que fazem essa parte do trabalho trabalham apenas quatro horas diárias no projeto. Mas, se abstrairmos esse fato, isso significaria que um lexicógrafo em tempo integral em um dia de trabalho de oito horas seria capaz de terminar cerca de 40 acepções e em uma semana de 40 horas cerca de 200 acepções, o que resulta em um total de cerca de 9.000 acepções por lexicógrafo em um ano com carga horária de 45 semanas.

O resultado é que os quatro lexicógrafos vinculados ao projeto, trabalhando por meio turno, concluíram um total de cerca de 40.000 registros (cartões) de março de 2014 a julho de 2016. Uma comparação poderia ser feita com o pequeno exército de lexicógrafos que está trabalhando no dicionário da Real Academia Española e que, acreditamos, conta com cerca de 20 pessoas. Se essa equipe trabalhasse com a ferramenta e o método descritos, seriam capazes de produzir cerca de 180.000 acepções por ano. Se trabalhassem durante três anos, passariam para mais de meio milhão de acepções e, passados apenas cinco anos e meio, o número atingiria um milhão de acepções, ou seja, o dicionário de espanhol mais completo já produzido. Em relação a isso, o desafio está feito! Basicamente, trata-se de se adaptar e explorar plenamente as novas tecnologias e técnicas colocadas à disposição da lexicografia. Os Dicionários On-line de Espanhol Valladolid-UVa fornecem um exemplo de como isso pode ser feito, embora não afirmemos que este é o único caminho que leva à Roma.

8. Competências e papel ativo do lexicógrafo

Apenas ferramentas, métodos e bases empíricas não constituem um dicionário por si só. Por mais avançada que seja a tecnologia, o mais importante na criação de dicionários ainda é o fator humano, através de lexicógrafos habilidosos, experientes e motivados. Dicionários, como Gudmann (2014: 31) afirma corretamente, “ainda são feitos por seres

humanos reais através de um processo criativo, sem que exista uma resposta absolutamente correta”.

Então, o que exatamente é exigido dos lexicógrafos que participam do projeto? Aqui, mais uma vez, é necessário fazer uma distinção entre conhecimento e habilidades. Isto significa, por um lado, que um projeto *on-line* moderno, como os Dicionários On-line de Espanhol Valladolid-UVa, não pode prosperar sem um gerente (lexicógrafo-chefe) que tenha um profundo conhecimento da teoria e da metodologia lexicográfica, bem como a capacidade de projetar um conceito de dicionário, escrever instruções, selecionar e treinar uma equipe de colaboradores e supervisionar o trabalho diário.

Por outro lado, também significa que o projeto precisa de uma equipe de lexicógrafos habilidosos, altamente produtivos e capazes de gerar dados lexicográficos com a qualidade desejada. Esses profissionais devem, acima de tudo, ter competências linguísticas em espanhol, ou seja, devem ser falantes nativos da língua espanhola. Além disso, eles também devem ter “um bom conhecimento do mundo em geral e sobre, pelo menos, uma área específica” (Bergenholtz 2013: 5). O conhecimento específico pode ser sobre linguística, mas também pode ser sobre outras disciplinas relevantes para o projeto. A esse respeito, Bergenholtz (2013) relata que a equipe de lexicógrafos que trabalha nos dicionários dinamarqueses *on-line* é composta por pessoas de estudos linguísticos, da matemática, da química, da biologia molecular, da física, das ciências jurídicas, da economia e da química. Quando os lexicógrafos espanhóis foram testados antes de serem empregados no projeto Valladolid-UVa, além da competência e do conhecimento da língua, tinham o dever de mostrar outras habilidades relevantes, como a capacidade de usar computadores, navegar na Internet, encontrar dados relevantes de acordo com as instruções e transformar esses dados em definições facilmente compreensíveis na língua espanhola.

No entanto, este é apenas o ponto de partida. Em algumas das ações listadas na Seção 5, fica claro que a seleção de itens de significado não é uma ciência exata com “uma resposta absolutamente correta”. Pelo contrário, um resultado bem-sucedido depende, em grande parte, do papel e das decisões ativas do lexicógrafo, que não se baseiam apenas nas competências e nos conhecimentos de linguagem, mas também na experiência. Pelo menos, este é o caso das seguintes ações:

- Ação 4: Quantas páginas devem ser ignoradas?
- Ação 6: Quais partes dos mini-textos são relevantes?
- Ação 8: Quantas páginas devem ser revisadas?
- Ação 9: Quando a quantidade de dados empíricos é suficiente?
- Ação 11: Quando o lexicógrafo está satisfeito com o processo?

As decisões tomadas nesses casos afetarão a qualidade do produto final e o próprio tempo para tomar a decisão terá consequências na produtividade. O desafio é, portanto, reduzir o tempo de tomada de decisão e elevar a qualidade das decisões. Para entender o que acontece, ou deveria acontecer, é útil referir-se aos “cinco estágios de aquisição de

habilidades” propostos por Dreyfus & Dreyfus (1986), que operam com os seguintes tipos de estágios de acordo com as habilidades: novato, iniciante avançado, competente, proficiente e experiente. Flyvbjerg (2001) resumiu as características dessas cinco etapas no processo de aprendizagem:

(1) Os novatos agem com base em elementos e regras independentes do contexto. (2) Iniciantes avançados também usam elementos situacionais, que aprenderam a identificar e interpretar com base em sua própria experiência em situações semelhantes. (3) Os competentes são caracterizados pela escolha de metas e planos como base para suas ações. Metas e planos são usados para estruturar e armazenar informações dependentes e independentes do contexto. (4) Os proficientes identificam problemas, metas e planos intuitivamente a partir da própria perspectiva baseada na experiência. A escolha intuitiva é verificada pela avaliação analítica antes da ação. (5) Finalmente, o comportamento dos experientes é intuitivo, holístico e sincrônico, compreendido na maneira como determinada situação libera um quadro de problema, meta, plano, decisão e ação em um instante e sem divisão em fases. Este é o nível da verdadeira expertise humana: os experientes são caracterizados por um desempenho fluente e sem esforço, livre de deliberações analíticas. (Flyvbjerg 2001: 20-21)

Flyvbjerg (2001: 21) acrescenta que o modelo acima contém um “salto qualitativo” dos três primeiros estágios para os estágios 4 e 5, e que “o salto implica um abandono do pensamento baseado em regras como a base mais importante para a ação, e sua substituição por contexto e intuição”. Em outra publicação, Dreyfus e Dreyfus (1992) enfatizam esse ponto:

Parece que os iniciantes fazem julgamentos usando regras e características rígidas, mas, com talento e muita experiência, o iniciante se torna um especialista que vê intuitivamente o que fazer sem precisar aplicar regras e fazer julgamentos. A tradição intelectualista resultou em uma descrição precisa do iniciante e do especialista diante de uma situação desconhecida, mas normalmente um especialista não deliberou os resultados. Ele não raciocina. Ele nem sequer age deliberadamente. Ele simplesmente faz espontaneamente o que normalmente funcionou e, portanto, normalmente funciona. (Dreyfus e Dreyfus 1992: 117)

Se esse modelo for transferido para a lexicografia, será mais fácil entender o que faz um bom lexicógrafo e o que deve ser levado em conta ao selecionar uma equipe de colaboradores para um projeto de dicionário.

No caso concreto do projeto Valladolid-UVa, quando o anúncio de emprego foi publicado, mais de 90 candidatos demonstraram interesse. Destes, 12 candidatos foram pré-selecionados com base em seus conhecimentos e competências documentados. Foi oferecido a eles (fevereiro de 2014) um curso de 30 horas ministrado por Pedro A. Fuertes-Olivera e Helene R. Gudmann, uma lexicógrafa habilidosa com experiência em dicionários de Internet dinamarqueses. O curso incluiu a introdução à lexicografia e ao projeto Valladolid-UVa, além de instruções sobre como coletar dados na Internet, escrever definições e preparar as categorias de dados restantes. Os 12 candidatos fizeram um teste no qual deveriam preencher um número de cartões no banco de dados com base nas instruções. Os quatro melhores profissionais foram selecionados e começaram a trabalhar por um período experimental de três meses - uma exigência da legislação espanhola.

No momento do teste, os quatro lexicógrafos que, posteriormente, conseguiram o emprego poderiam ser caracterizados como *novatos* de acordo com o modelo de Dreyfus. No período experimental de três meses, esperava-se que eles se desenvolvessem relativamente rápido para *iniciantes avançados* e depois para *profissionais competentes*, uma transformação que se refletiria no aumento da produtividade e da qualidade do trabalho lexicográfico. Se não fosse o caso, o contrato seria cancelado. No entanto, após três meses, a tendência é que os profissionais sejam ainda dependentes das instruções lexicográficas (regras) e dos objetivos e planos escolhidos como base para suas ações e decisões, embora alguns deles possam começar a usar sua intuição com base em experiências anteriores. Se esse for o caso, eles saltarão para o estágio quatro, o do profissional *proficiente*, que é o nível mínimo que deveria ser esperado de todos os lexicógrafos que participam de projetos de tal magnitude e importância, como o dos Dicionários On-line de Espanhol Valladolid-UVa.

O quinto estágio no modelo de Dreyfus é o nível de *experiente*, o qual só é alcançado por uma porção de lexicógrafos praticantes. Depende, acima de tudo, da experiência, mas não pode ser alcançado sem talento, o qual deve ser identificado no período experimental. Nesse estágio, o lexicógrafo simplesmente age espontaneamente sem deliberar e, portanto, “mesmo os melhores lexicógrafos, quando pressionados, nunca conseguem explicar o que estão fazendo ou por qual razão” (Wierzbicka 1985: 5). A lexicografia tornou-se puramente artesanal. No entanto, isso certamente não implica que “a lexicografia não tem fundamento teórico”, como também afirma Wierzbicka, ou que o processo de compilação lexicográfica não se baseia em regras. Significa, na verdade, que essas regras foram completamente internalizadas e integradas com a intuição baseada na experiência, em um desempenho fluente, holístico e que exige pouco esforço, em que o lexicógrafo, como qualquer outra pessoa que atua nesse nível, não sabe explicar, exatamente, o que está fazendo. A teoria lexicográfica e as instruções (regras) ainda se

fazem presentes, como pano de fundo, sendo uma importante e necessária instrumentária para transmitir conhecimentos e habilidades para futuros lexicógrafos.

Atualmente, os lexicógrafos que participam do projeto na Universidade de Valladolid podem ser caracterizados como executores proficientes ou experientes, conforme definido no modelo Dreyfus. Isto implica que eles agora são capazes de tomar decisões rápidas, qualificadas e intuitivas para agir em situações como as discutidas acima (Ações 4, 6, 8, 9 e 11), bem como em todas as outras situações que podem estar relacionadas ao processo de compilação. Nesse sentido, a intuição humana baseada na experiência é um importante fator de produção sem o qual o sucesso em um projeto lexicográfico seria impossível.

Essa lógica pode ser resumida da seguinte forma: em primeiro lugar, a tecnologia e os métodos necessários para trabalhar diretamente com a Internet como um *corpus* exigem seres humanos qualificados, instruídos e talentosos, motivados a ter um desempenho extraordinário; e, em segundo lugar, essas características devem ser percebidas pelo gerente de projeto o mais rapidamente possível - uma habilidade que também exige experiência e talento. Essa é, pelo menos, a experiência de trabalho atual dos Dicionários On-line de Espanhol Valladolid-UVa.

9. Conclusões

Os *corpora* foram introduzidos na década de 1960, enquanto a Internet, como fenômeno generalizado, foi disseminada até três décadas depois, na década de 1990. É surpreendente que até agora a lexicografia tenha feito mais uso de uma tecnologia antiga do que de uma mais recente. A questão é se essa seria a hora de explorar as possibilidades lexicográficas da Internet. A experiência dos Dicionários Valladolid-Uva indica claramente que o tempo é mais do que oportuno. Isso mostra que os lexicógrafos habilidosos e bem treinados que trabalham com as ferramentas e os métodos certos são perfeitamente capazes de lidar com as eventuais desvantagens ao usar a Internet diretamente como um *corpus* em vez dos tradicionais *corpora* de texto. Isso não significa, obviamente, que os *corpora* não tenham mais relevância para a lexicografia, pois ainda têm um papel importante a desempenhar relacionado a diversas tarefas. Significa, acima de tudo, que a lexicografia, para enfrentar a atual crise, precisa estar no ambiente virtual não apenas para apresentar seus produtos, mas também para deixá-los com a qualidade e a produtividade necessárias.

Neste artigo, discutimos um conjunto de dicionários *on-line* de espanhol, uma das línguas mais faladas do mundo, com mais de quatrocentos milhões de falantes nativos. É evidente que o número de textos em espanhol disponíveis na Internet é enorme. No entanto, de acordo com a experiência dos dicionários *on-line* dinamarqueses, é perfeitamente possível usar a mesma tecnologia e metodologia ao trabalhar com uma língua menos falada, com apenas cinco milhões de falantes maternos. Em relação a isso, o possível problema não é tanto o número de falantes, mas sim a penetração e o uso generalizado da Internet dentro de certa comunidade de fala. Isso sugere que pode haver

algumas línguas africanas com relativamente poucos falantes, em que a coleta de textos da Internet ainda não é grande o suficiente para compilar dicionários, conforme aqui descrito – mas elas serão a exceção à regra. Na maioria dos casos, a quantidade de textos da Internet já é suficiente, ou será no futuro próximo. A Internet está aqui para ficar, pelo menos por alguns anos, e seria um grande erro não começar a explorar hoje suas possibilidades lexicográficas⁸.

Referências

Dicionários

BERGENHOLTZ, H. Ed. 2016. *De Danske Netordbøger*. Odense: Ordbogen.com.

FUERTES-OLIVERA, P.A.; H. BERGENHOLTZ (Eds.) in collaboration with M.Á. Sastre Ruano, E. Álvarez Ramos, M. Fonseca Hernández, M.J. López Carrero, Á. Prieto Salvador and O. Saldaña. *Diccionarios en Línea de Español “Universidad de Valladolid”*. Hamburg: Lemma.com. (Under construction).

MOLINER, M. 2007. *Diccionario de Uso del Español*. Third Edition. Madrid: Gredos.

Real Academia Española. 2014. *Diccionario de la Lengua Española*. 23rd Edition. Madrid: Espasa.

SÁNCHEZ PÉREZ, A. Ed. 2001. *Gran Diccionario de Uso del Español Actual*. Madrid: Sociedad General Española de Librería.

SECO, M.; O. ANDRÉS; G. RAMOS. 2011. *Diccionario del Español Actual*. Madrid: Aguilar.

Leituras adicionais

ATKINS, B.T.S.; M. RUNDELL. 2008. *The Oxford Guide to Practical Lexicography*. Oxford, New York: Oxford University Press.

BERGENHOLTZ, H. 1996. Korpusbaseret Leksikografi. *LexicoNordica* 3: 1-15.

BERGENHOLTZ, H. 2013. The role of Linguists in Planning and Making Dictionaries in Modern Information Society. D. Kwary, N. Wulan and L. Musyahda. Eds. *Lexicography and Dictionaries in the Information Age. Selected papers from the 8th ASIALEX Conference*: 1-10. Surabaya: Airlangga University Press.

BERGENHOLTZ, H. ; H. AGERBO. 2014. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes* 52: 125-139.

⁸ N.A.: Agradecemos ao Ministério da Economia e Competitividade espanhol pelo financiamento do projeto “La Teoría Funcional de la Lexicografía: Diseño y Construcción de Diccionarios de Internet” (Ref. FFI2014-52462-P).

- BERGENHOLTZ, H.; S. TARP. Eds. 1995. *Manual of Specialised Lexicography*. Amsterdam, Philadelphia: John Benjamins.
- DREYFUS, H.; S. DREYFUS. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- DREYFUS, H.; S. DREYFUS. 1992. What is Moral Maturity? Towards a Phenomenology of Ethical Expertise. J. Ogilvy. Ed. *Revisioning Philosophy*: 111-131. Albany: State University of New York Press.
- FLYVBJERG, B. 2001. *Making Social Science Matter. Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge: Cambridge University Press.
- FRANCIS, W.N. 1979. Problems of Assembling and Computerizing Large Corpora. H. Bergenholtz and B. Schaefer. Eds. *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*: 110-123. Königstein/Ts.: Scriptor.
- FUERTES-OLIVERA, P.A. 2012. Lexicography and the Internet as a (Re-)source. *Lexicographica* 28: 49-70.
- FUERTES-OLIVERA, P.A.; H. BERGENHOLTZ. 2015. Los Diccionarios en Línea de Español “Universidad de Valladolid”. *Estudios de Lexicografía* 4: 71-98.
- FUERTES-OLIVERA, P.A.; S. TARP. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin, Boston: De Gruyter.
- GOUWS, R.H.; S. TARP. 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography*. 30 (4), 389-415.
- GUDMANN, H.R. 2014. *Betydningshuller i Spanske Ordbøger. En Undersøgelse af Betydningsenheder i Spanske Monolingvale Almene Receptionsordbøger*. Master Thesis. Aarhus: Aarhus University, Department of Business Communication.
- GUDMANN, H.R. 2015. Lagunas de Significado en los Diccionarios Españoles. *Estudios de Lexicografía* 4: 161-184.
- HANKS, P. 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25 (4): 398-436.
- KILGARRIFF, A. 1997. I Don’t Believe in Word Senses. *Computers and the Humanities* 2 (31): 91-113.
- KILGARRIFF, A. 2012. [Review of] Pedro A. Fuertes- Olivera/Henning Bergenholtz (Eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. *Kernerman Dictionary News*, July 2012: 26-29.
- KILGARRIFF, A.; G. GREFENSTETTE. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29: 333-347.
- LEES, R. 1962. Oral contribution. Quoted by Francis (1979): 110.

RUNDELL, M. 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.

SINCLAIR, J.M. 1997. Introduction. *Collins Cobuild English Language Dictionary*. London: HarperCollins Publishers, xv-xxi.

TARP, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.

_____. 2015. Structures in the Communication between Lexicographer and Programmer: Database and Interface. *Lexicographica* 31: 17-46.

_____. 2016. Excesos en el Uso de Corpus en la Lexicografía: «Pesca» de Términos y Definiciones. *Revista de Lexicografía* 21: 145-163.

TARP, S.; P.A. FUERTES-OLIVERA. 2016. Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva. *Lexikos* 26: 273-295.

WIERZBIECKA, A. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

XUE, M.; S. TARP. 2016. Corpus-based, Corpus-driven or Corpus-assisted lexicography? The Limited Usefulness of Corpora in Defining Specialised Terms. *Lexicographical Studies*, 4: 1-11.

Como citar este texto (ABNT):

TARP, S.; FUERTES-OLIVEIRA, P.A. Tradução de Luísa Rabaldo. Métodos e técnicas para usar a Internet diretamente como *corpus*: o caso dos Dicionários on-line de Espanhol Valladolid-UVa. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 10-32, 2018.

DICIONÁRIOS SENSÍVEIS AO CONTEXTO E INTEGRADOS A ASSISTENTES DE ESCRITA EM L2: NOVOS DESAFIOS PARA A LEXICOGRAFIA¹

Sven Tarp²

Kasper Fisker³

Peter Sepstrup⁴

Tradução: Guilherme S. de Oliveira⁵

Supervisão e revisão: Márcia Moura da Silva⁶

*“Levantar novas questões, novas possibilidades, considerar velhas questões de um novo ângulo, isso requer imaginação criativa e marca avanços reais na ciência.”*⁷

(EINSTEIN & INFELD, 1938, p. 92)

Resumo: Os dicionários estão cada vez mais integrados a outras ferramentas criadas para ajudar na leitura, escrita e tradução de textos. O Write Assistant é um novo tipo de ferramenta que ajuda as pessoas a escreverem em uma segunda língua. Ele se alimenta de um *big data* colhido de *corpora* e dicionários digitais. Este artigo discute a filosofia por trás das técnicas aplicadas, sua base empírica e funcionalidade, assim como a que ponto ele realmente ajuda os usuários. Mostra também como a ferramenta possibilita a redução, e até mesmo a eliminação, de algumas fases do processo tradicional de pesquisa de informações e permite ao usuário manter o foco na mensagem a ser escrita sem a necessidade de procurar por informação em recursos externos. O estudo ainda aponta como a tecnologia subjacente dá origem a um novo tipo de dicionário que é sensível ao contexto e oferece um serviço mais personalizado ao usuário. Por fim, ele indica que os novos dicionários precisam ser conceitualmente adaptados à ferramenta específica para otimizar o serviço. Tudo isso impõe novos desafios para a Lexicografia.

Palavras-chave: assistente de escrita; ferramenta de informação; dicionários integrados; escrita em L2; processo de pesquisa de informação; recursos empíricos; *corpora*; modelo de língua; dicionários sensíveis ao contexto; conceito de dicionário.

¹ N.A.: Este artigo é uma adaptação de um texto publicado em *Lexikos* 27 (cf. TARP, FISKER, SEPSTRUP, 2017). N.E.: A presente tradução foi autorizada para ser publicada em português pelo autor Sven Tarp.

² Departamento de Africâner e Dinamarquês, Universidade de Stellenbosch, África do Sul e Centro de Lexicografia, Universidade Aarhus, Dinamarca (st@cc.au.dk).

³ Ordbogen A/S, Odense, Dinamarca (kfi@ordbogen.com).

⁴ Ordbogen A/S, Odense, Dinamarca (pse@ordbogen.com).

⁵ Bacharelado em Letras – Português/Inglês, Instituto de Letras, UFRGS.

⁶ Professora do Departamento de Línguas Modernas, Instituto de Letras, UFRGS.

⁷ Todas as citações diretas foram traduzidas pelo tradutor do presente artigo.

1. Introdução

Com o passar dos anos, dicionários têm sido desenvolvidos e compilados para atingir uma grande variedade de necessidades humanas detectadas na sociedade. Uma dessas muitas funções é ajudar usuários que escrevem textos em língua estrangeira. Como Chon (2009) relata, pessoas que praticam essa atividade consultam frequentemente uma combinação de dicionários L1-L2, L2 e L2-L1 quando se deparam com diferentes tipos de problemas e necessitam informações para resolvê-los. A tradição de consultar uma combinação de dicionários monolíngues e bilíngues foi passada do mundo impresso ao mundo digital; é surpreendente que relativamente poucos dicionários, como o *WorldReference*, foram concebidos como um pacote integrado que pode suprir as necessidades de ambos usuários, monolíngues ou bilíngues, em conexão com a produção textual em L2. Apesar dessa classe de dicionários indubitavelmente representar um importante passo para a Lexicografia, o momento parece oportuno para dar um passo ainda maior em termos de prestação de serviço mais personalizada e eficiente. A tecnologia necessária para isso já existe. É meramente uma questão de aproveitar as tecnologias da informação por completo.

Um dos vários desafios atuais é repensar todo o processo que as pessoas tradicionalmente seguem quando procuram por informações. Esse processo é mais ou menos descrito por Bergholtz *et al.* (2015):

1. alguém se depara com a falta de informação em uma situação específica;
2. percebe essa falta;
3. decide agir e consulta um detentor de informação, por exemplo, um dicionário;
4. realiza a consulta;
5. avalia os resultados, por exemplo, se são satisfatórios;
6. retorna à situação onde a informação é necessária originalmente;
7. e usa a informação coletada para resolver o problema que criou essa necessidade.

Esse processo é bastante complexo e exigente. Em uma perspectiva lexicográfica, é necessário que, dentre outras coisas, a pessoa que se depara com essa falta a) esteja ciente dela; b) possa identificar suas características; c) tenha um dicionário à mão e saiba como consultá-lo, encontrando dados relevantes; d) seja capaz de retirar a informação necessária desses dados; e) esteja suficientemente preparada para avaliar e retirar a informação; e f) saiba como aplicá-la. Além disso, o processo é, de qualquer forma, *demorado e incômodo*, porque ele retira o foco da atividade principal que a pessoa está realizando, especialmente se é uma atividade comunicativa, como produção textual em uma língua não nativa na qual a mensagem a ser transmitida é a questão central. Nesi (2015, p. 584) observa corretamente que “pessoas tipicamente consultam mapas, enciclopédias e dicionários enquanto estão fazendo outra coisa”. Nesse caso, qualquer consulta de uma fonte de

informação externa inevitavelmente representa uma interrupção da atividade em questão. Pode-se supor que a maioria dos usuários desses recursos queira apenas voltar o mais rápido possível ao que estava fazendo para manter o foco.

Como uma complicação adicional, escritores procurando por assistência em um dos muitos dicionários disponíveis na internet correm o risco de encontrar tentações imprevistas. Muitos usuários do século XXI, especialmente os jovens, esperam que essas consultas sejam gratuitas. Editoras do mundo todo vêm tendo problemas para encontrar um modelo de negócio economicamente viável, um desafio que tem levado algumas delas a dependerem de propagandas como uma estratégia de sobrevivência. Isso pode ser bom para os negócios, mas não necessariamente para os usuários. Pesquisas realizadas sobre o fenômeno indicam que pelo menos 5% de todos os visitantes de *websites* que utilizam propagandas sentem-se tentados pelos *banners* coloridos e clicam para conseguir mais informações, vide ROBINSON *et al.* (2007). Mesmo aqueles que sabem que nem tudo que reluz é ouro ainda podem distrair-se com essas muitas tentações. Quando finalmente retornam à tarefa que estavam realizando, eles provavelmente terão perdido o foco e talvez até esquecido o porquê de terem começado a consulta.

Por isso, da perspectiva do usuário, o desafio é desenvolver uma ferramenta de consulta que torne possível reduzir ou até mesmo eliminar algumas das fases e passos mencionados acima para diminuir o tempo de consulta geral, mantendo o fluxo de escrita, e evitando que a pessoa em questão perca a concentração. Em relação a isso, as palavras-chave são *tempo*, *fluxo* e *foco*, às quais deveria ser adicionada *qualidade* da informação fornecida.

Nas seções seguintes, analisaremos uma ferramenta que tem a intenção de aceitar o desafio, o Write Assistant, desenvolvido pela Ordbogen A/S, uma empresa dinamarquesa de TI que trabalha com serviços relacionados à linguagem e dicionários on-line, gerais e especializados. Discutiremos a filosofia por trás da ferramenta, as técnicas aplicadas, sua base empírica e funcionalidade, a amplitude da ajuda prestada ao usuário, assim como os novos desafios impostos à Lexicografia. Entretanto, antes disso, iremos examinar a relação entre Lexicografia e Tecnologia bem como os requisitos para auxiliar a produção textual em L2.

2. Lexicografia e tecnologia: avanços recentes

Historicamente, há uma íntima relação entre Lexicografia e Tecnologia, como foi mostrado por Hanks (2011), Rundell e Kilgarrieff (2011), e outros. A tecnologia tem grande influência no *design*, na compilação, na apresentação, na distribuição, na disponibilidade e no uso dos dicionários. Novas invenções disruptivas como as tecnologias da prensa e do computador têm levado a genuínas mudanças de paradigma com consequências consideráveis para a disciplina. Fuertes-Olivera (2016) descreve a situação atual da Lexicografia como uma “explosão Cambriana”. Apesar do progresso inegável que pode ser observado, a Lexicografia ainda está, em muitos aspectos, em processo de adaptação completa ao que as tecnologias de informação disponibilizam. A oferta de uma resposta

lexicográfica para as crescentes demandas sociais para um serviço mais personalizado está entre os principais problemas estratégicos; vide RUNDELL (2010) e TARP (2011).

Apesar de a Lexicografia poder ser vista como uma disciplina em si, ela é tradicionalmente caracterizada por seu espírito interdisciplinar; vide NIELSEN (2017) e TARP (2017). Essa situação é ainda mais nítida hoje, onde nenhum projeto lexicográfico sério pode ser executado sem o conhecimento combinado da Lexicografia, da Tecnologia da informação, da Linguística e de outras disciplinas relevantes ao projeto. Nenhum especialista de qualquer uma dessas áreas poderia produzir um trabalho lexicográfico de alta qualidade sozinho. Qualquer um que tome a iniciativa tem que atravessar a fronteira disciplinar e incluir conhecimentos e habilidades das outras áreas relevantes. As ferramentas de informação integradas discutidas abaixo jamais veriam a luz do dia sem os esforços combinados de especialistas em TI, lexicógrafos, linguistas de corpus, *web designers* etc. Assim, o papel central da Lexicografia é determinado pela longa experiência e habilidade de especificar as necessidades do usuário, por definir o dado lexicográfico correspondente e por estabelecer o melhor meio de acessar esses dados. Esse papel vai muito além da Lexicografia tradicional e de construção de dicionários como previsto por Tarp (2009).

Um dos desenvolvimentos recentes mais promissores é o crescente desafio do dicionário tradicional, impresso ou digital, e a integração gradual dos produtos lexicográficos com outras ferramentas de informação digitais desenvolvidas para auxiliar na leitura, escrita e tradução de textos; vide, por exemplo, VERLINDE *et al.* (2010), PAQUOT (2012), VERLINDE e PEETERS (2012), e GRANGER e PAQUOT (2015). Outro desenvolvimento importante afeta o próprio conceito de consulta, pois agora um grande número de pesquisas é feito automaticamente; fenômeno definido por Tarp (2008, p. 123) como “pesquisa passiva” em contraste com a “pesquisa ativa” realizada pelos próprios usuários. Essas consultas “passivas” acontecem frequentemente sem que o usuário saiba que um dicionário está sendo consultado. Apesar de não haver estatísticas confiáveis, a maioria das consultas em dicionários hoje é, provavelmente, feita em algum dos vários tipos de ferramenta integrada de informação e até automáticas.

Essas duas inovações apontam na direção certa. A integração de dicionários em outras ferramentas permite a redução dos “custos de informação” em termos de tempo gasto na consulta e em processamento e aplicação da informação encontrada; vide NIELSEN (2008). Além disso, a introdução da “pesquisa passiva” nos dicionários e outras fontes de informação tende a neutralizar o conhecido problema de que “muitos usuários se cansam de consultar dicionários sempre que encontram uma palavra desconhecida” (VERLINDE; PEETERS, 2012, p. 158). Ambas inovações representam um passo em direção a um serviço melhorado que deixa os usuários dessas ferramentas com mais tempo para focar na atividade primária, sendo ela ler, escrever ou traduzir.

Quando se trata de auxílio na escrita, as ferramentas criadas para prover esse serviço podem ser concebidas com funcionalidades diferentes dependendo da filosofia sustentada. A *Interactive Language Toolbox* de holandês-francês-inglês, por exemplo, foi

primariamente concebida como uma ferramenta de aprendizado de língua e portanto “não corrige o texto enviado”, mas “apenas identifica padrões sintáticos e lexicais que podem conter erros” para assim encorajar o usuário a “refletir criticamente sua escrita” (VERLINDE, 2011, p. 282-283).

Com esse plano de fundo, uma distinção pode ser feita entre *assistentes de escrita de detecção, correção e previsão*. O primeiro se refere aos vários tipos de verificador de ortografia e gramática, mas esforços têm sido feitos para desenvolver uma ferramenta que possa verificar outras categorias linguísticas como colocações; vide WANNER *et al.* (2013). A vantagem principal de um *assistente de escrita* de detecção é que eles podem chamar a atenção do usuário a problemas (e necessidades) dos quais ele pode não estar ciente. Por outro lado, os *assistentes de escrita* de correção, por exemplo os fornecidos pela Microsoft, não apenas detectam possíveis erros no textos já escrito, mas também oferecem soluções alternativas. E, finalmente, as ferramentas de previsão intervêm diretamente no processo de escrita com sugestões, ou como completar uma palavra quando a primeira letra é digitada, ou qual palavra poderia ser a próxima na frase. Muitas pessoas conhecem essas ferramentas interativas e previsoras por causa de seus smartphones e tablets, nos quais elas têm o potencial de agilizar o processo geralmente realizado com apenas um ou dois dedos.

3. Necessidades de informação na produção de textos em L2

Há uma literatura relativamente vasta que lida com as necessidades que as pessoas podem encontrar quando escrevem em uma segunda língua, assim como com as respostas que os dicionários deveriam fornecer a essas necessidades; vide, por exemplo, RUNDELL (1999), TARP (2004), BOGAARDS (2005), CHON (2009), e LEW (2016). Essa literatura contém ideias valiosas que serão levadas em consideração nas próximas reflexões. Entre os lexicógrafos mencionados, há consenso de que nem dicionários monolíngues, nem bilíngues são capazes de suprir sozinhos todas as necessidades dos usuários. Apenas uma combinação de dicionários pode realizar isso. Mas que combinação? E quais dados lexicográficos esses dicionários deveriam oferecer para auxiliar os usuários na produção de texto em L2? Uma tentativa de responder essas questões será feita nesta seção. A discussão será baseada na ideia de que as necessidades do usuário são primeiramente determinadas pela situação na qual elas ocorrem e então moldadas pelas características relevantes da pessoa que se depara com elas (FUERTES-OLIVERA; TARP, 2014, p. 48-57). O ponto de início para determinar essas necessidades, portanto, é uma análise da situação de escrita em L2 e dos diferentes tipos de problema que um escritor pode encontrar nessa situação.

Se excluirmos os erros já cometidos dos quais o autor não está ciente, os problemas mais comuns e necessidade de informação correspondentes estão provavelmente incluídos no Top 10 abaixo, no qual “palavra” é usada como um termo genérico que inclui tanto as palavras simples como as compostas:

Quadro 1 – Problemas e necessidades de informação frequentes na escrita em L2

Classe de problema		→	Necessidade de informação
1	O autor conhece uma palavra em L2, mas não consegue se lembrar dela.	→	Equivalente L2 fornecida em uma solução L1-L2
2	O autor não conhece a palavra a ser usada em L2.	→	Equivalente L2 fornecida em uma solução L1-L2 com diferenciação de sentido
3	O autor não conhece a colocação a ser usado em L2.	→	Colocação L2 fornecida em uma solução L1-L2
4	O autor conhece uma palavra em L2, mas não tem certeza se ela pode ser usada com o sentido concreto.	→	Explicações
5	O autor conhece uma palavra em L2, mas não tem certeza se ela pode ser usada no contexto.	→	Informações estilísticas, pragmáticas, culturais ou político-linguísticas
6	O autor conhece uma palavra em L2, mas não tem certeza como escrevê-la.	→	Informação sobre ortografia (ou autocorreção)
7	O autor conhece uma palavra em L2, mas não tem certeza de como ela é flexionada.	→	Informação sobre as formas flexionadas
8	O autor conhece uma palavra em L2, mas não tem certeza de como combiná-la a outras palavras.	→	Informação explícita ou implícita sobre suas propriedades sintáticas.
9	O autor conhece uma palavra em L2, mas não tem certeza de como construir uma colocação com ela.	→	Colocações com a palavra em L2 em questão
10	O autor conhece uma palavra em L2, mas quer variar o vocabulário e usar outra palavra.	→	Sinônimos e/ou antônimos

Fonte: Elaborado pelo autor

É óbvio que o Top 10 dado acima não representa todos os problemas complexos e necessidades relativos à escrita em L2, como é o caso dos problemas relacionados à fraseologia e aos vários tipos de expressões cristalizadas. Entretanto, representa indubitavelmente alguns dos pontos mais relevantes, apesar de eles não poderem ser vistos isolados do par real de línguas em questão. Gênero, por exemplo, não é relevante em inglês, mas certamente é em línguas como alemão, espanhol e italiano. A isso, podem ser

adicionadas outras classes de problema com as quais as pessoas podem se deparar enquanto escrevem textos em uma outra L2, especialmente nas línguas de fora da família Indo-Europeia.

É importante notar a expressão “não tem certeza” usada em relação às várias classes de problema (4-9). Essa expressão cobre duas variantes do mesmo problema, sendo 1) o autor não conhece o item linguístico em questão, e 2) o autor conhece o item, mas não consegue se lembrar dele, portanto precisa de um lembrete, experiência muito frequente para estudantes de L2 que tipicamente aprendem vocabulário e gramática muito mais passivamente do que ativamente.

Como pode ser observado, apenas problemas das três primeiras classes listadas acima requerem uma solução L1-L2 visto que os problemas remanescentes requerem uma solução baseada em L2, por exemplo, L2 monolíngue ou L2-L1 bilíngue. Autores inexperientes e estudantes em nível iniciante podem encontrar mais problemas nas primeiras três classes do que um autor experiente e um estudante avançado, mas todos eles podem de vez em quando enfrentar problemas de todas as dez classes. A separação tradicional de estudantes em iniciante, intermediário e avançado, portanto, não parece ser relevante, embora possa ser argumentado que crianças com problemas na escrita em L2 podem precisar de informações menos complexas do que um autor adulto; vide NOMDEDEU e TARP (2017).

4. Write assistant

A seguir, discutiremos sobre o Write Assistant desenvolvido pela Ordbogen A/S. Essa ferramenta é destinada a dar respostas instantâneas e de alta tecnologia aos problemas referidos na seção anterior. Assim sendo, não foi desenhado como um programa de processamento de texto independente, mas como um *aplicativo* que atualmente só pode ser usado com os programas do pacote Office da Microsoft, apesar de não haver nada que o impeça de ser adaptado a outros tipos de *software* quando necessário. Isso significa que o autor que usa o aplicativo simultaneamente pode beneficiar-se das vantagens do pacote Office, incluindo verificação de ortografia e gramática. O Write Assistant não é, portanto, desenhado principalmente com funções de detecção e correção, mas sobretudo com a combinação das funções de *previsão* e *tradução* como veremos abaixo.

Depois de baixado, o aplicativo só pode ser ativado quando o Word é a janela ativa em primeiro plano. Ele aparecerá como um ícone no centro do documento Word, que deve ser movido para fora do documento para ser ativado. A partir daí, ele funcionará em todos os programas do Office, incluindo Outlook, PowerPoint, Explorer etc. Na versão atual (junho 2017), o Write Assistant só está disponível para falantes nativos de dinamarquês que escrevem em inglês, mas a tecnologia desenvolvida permite a incorporação de qualquer par de línguas dentro de poucas semanas, desde que os recursos empíricos certos estejam disponíveis (veja abaixo). É previsto que a ferramenta esteja disponível em dezenas de línguas em um futuro próximo, juntamente com uma série de versões especificamente adaptadas à terminologia de companhias reais. A descrição e a

análise da ferramenta – incluindo sua filosofia, base empírica, modelo de língua, funcionalidade, utilidade, possibilidades, assim como limitações – são, portanto, relevantes para além da fronteira dinamarquesa.

4.1 Filosofia subjacente e base empírica

O Write Assistant se distingue dos programas de tradução automática que oferecem soluções mais ou menos adequadas, assim como dos dicionários e *corpora* que precisam da decisão ativa do usuário para serem consultados. A ferramenta não é um “loja de departamento” (BOWKER, 2012, p. 381), mas um *serviço por encomenda*. Não é “um conjunto de componentes que os clientes podem misturar e combinar de acordo com as necessidades” (RUNDELL, 2007, p. 50), mas um conjunto de componentes que a ferramenta utiliza para prover um *serviço customizado* ao usuário, de acordo com as prováveis necessidades de cada caso.

O Write Assistant não reivindica atender a todas as necessidades de informação que ocorrem durante o processo de escrita em L2. O aplicativo não foi desenhado para lidar com *qual* conhecimento é transmitido durante esse processo, mas apenas com *como* ele é transmitido. Se estudantes, por exemplo, precisam de informações sobre a relação próxima entre Johann Wolfgang von Goethe e Alexander von Humboldt para escreverem um texto sobre esse assunto, eles terão de pesquisar sobre isso em outro lugar. Mesmo assim, e como o nome sugere, o Write Assistant não foi concebido para oferecer soluções para as necessidades comunicativas de escritores, mas “apenas” auxiliá-los. Ainda é esperado dos usuários desse produto terem um papel decisivo, além de terem completa responsabilidade pela qualidade do texto final em L2. Como um aplicativo integrado e interativo, o propósito principal do Write Assistant é oferecer assistência instantânea, que os usuários podem ignorar ou aceitar sem desperdiçar tempo consultando recursos de informação externos, sem muitas interrupções no processo de escrita, e sem perder concentração e foco.

O Write Assistant se baseia em um *big data* criado, principalmente, a partir de dois recursos empíricos: um grande *corpus* L2; e um ou mais dicionários digitais, dentre eles um L1-L2, no mínimo, e outro dicionário monolíngue em L2. A razão pela qual o aplicativo se alimenta em *corpus* existente em vez de usar dados diretamente da internet tem a ver com a qualidade desses dados. Apesar de ser continuamente atualizado e muito maior do que qualquer outro *corpus*, a internet contém muito “ruído” na forma de textos não editados e palavras mal escritas. O Write Assistant precisa de dados de alta qualidade para oferecer um serviço de alta qualidade, e portanto até *corpora* existentes precisam ser mais “limpos” e itens com menos de dez ocorrências precisam ser deletados, como um meio de evitar o maior número possível de erros de ortografia.

Como os *corpora*, os dicionários a serem usados pelo Write Assistant como recurso de dados têm que estar disponíveis em uma plataforma digital, preferivelmente na forma dos conhecidos Modelos T da Ford que têm sido criados desde o início como mídias eletrônicas (TARP, 2011, p. 60). Na primeira edição a ser publicada, as pesquisas serão

feitas no dicionário bilíngue online dinamarquês-inglês/inglês-dinamarquês já gerenciados pela Ordbogen A/S e amplamente utilizado por seus assinantes. Desse modo, a edição de lançamento representa um bom ponto de partida, apesar da incorporação de um dicionário monolíngue de inglês também estar sendo considerado. Entretanto, e como será discutido abaixo, não se pode excluir que uma futura necessidade de otimização da qualidade do Write Assistant possa requerer a criação de produtos lexicográficos novos, ou parcialmente novos, que tenham seus dados adaptados a requisições muito específicas do aplicativo.

Resumindo, pode-se dizer que, à parte da tecnologia desenvolvida como uma precondição indispensável para qualquer sucesso, a qualidade do Write Assistant depende acima de tudo da qualidade dos *corpora* e dos dicionários usados como sua base empírica. Nesse ponto, tudo que é preciso para adaptar o aplicativo a qualquer par de línguas é a existência de um *big data* no formato de *corpora* e dicionários digitais. Se essa base empírica estiver à mão, o Write Assistant pode se tornar disponível em qualquer língua em um período de tempo muito curto.

4.2 Modelo de língua

Na versão beta atual, o Write Assistant faz pesquisas tanto em *modelo de língua* em inglês quanto no dicionário bilíngue dinamarquês-inglês/inglês-dinamarquês mencionado acima, por isso não pesquisa diretamente no *corpus* monolíngue do inglês. O modelo de língua é o resultado de uma análise e reconstrução automáticas das palavras contidas no *corpus*. Ele foi criado para receber *tuplas* de até quatro palavras e dar nota a cada uma. Exemplificaremos isso com a sequência “*I was drawn forward with the prospect of employment*” (Eu fui atraído pela perspectiva de conseguir um emprego) tirado do *The Plan of an English Dictionary* (1747), de Samuel Johnson. Essa sequência consiste de um total de seis tuplas de quatro palavras:

“*I was drawn forward*” (Eu fui atraído pela)

“*was drawn forward with*” (fui atraído pela perspectiva)

“*drawn forward with the*” (atraído pela perspectiva de)

“*forward with the prospect*” (pela perspectiva de conseguir)

“*with the prospect of*” (perspectiva de conseguir um)

“*the prospect of employment*” (de conseguir um emprego)

Baseado na frequência com que elas ocorrem no *corpus*, cada uma dessas seis tuplas receberá uma nota que indica a probabilidade do conjunto de quatro palavras aparecerem juntas na língua inglesa. O modelo de língua baseado nesses princípios tem três funções: 1) a finalização de palavras, 2) a previsão da próxima palavra na sentença, e 3) a priorização de candidatos à tradução.

A *finalização de palavras* será ilustrada com um exemplo real. Se o autor, por exemplo, digitou “*The man was married to a beautiful wo*” (O homem era casado com uma bela mu), o modelo destacará “*to a beautiful*” (com uma bela) e “*wo*”. Essas duas linhas são

chamadas de *contexto* e *prefixo*, respectivamente, onde a segunda não deve ser confundida com o termo linguístico tradicional. Por meio de uma lista classificada de palavras contidas no modelo de língua, o Write Assistant agora pesquisa todas as palavras com o prefixo “*wo*”. Elas poderiam ser *work* (trabalho), *woman* (mulher), *world* (mundo), *word* (palavra) etc. Essas possibilidades de “*wo*” serão adicionadas ao contexto “*to a beautiful*”, o que resulta nas seguintes tuplas de quatro palavras:

“ <i>to</i>	<i>a</i>	<i>beautiful</i>	<i>work</i> ”	(com	um	belo	trabalho)
“ <i>to</i>	<i>a</i>	<i>beautiful</i>	<i>woman</i> ”	(com	uma	bela	mulher)
“ <i>to</i>	<i>a</i>	<i>beautiful</i>	<i>world</i> ”	(com	um	belo	mundo)
“ <i>to</i>	<i>a</i>	<i>beautiful</i>	<i>word</i> ”	(com uma bela palavra)			
etc.							

A cada uma dessas tuplas será dada uma nota do modelo proporcional ao número de ocorrências no *corpus*. As finalizações nas dez tuplas que receberem a nota mais alta serão selecionadas e apresentadas ao usuário em uma ordem de frequência. No exemplo discutido aqui, é de se esperar que a finalização mais provável de “*wo*” seja “*woman*”. Entretanto, isso não deve ser entendido como uma solução final, mas como uma sugestão, já que o autor pode querer expressar algo diferente (por exemplo, dedicando um poema a um belo mundo, “*a beautiful world*”) e portanto tem que participar ativamente do processo e se responsabilizar pela palavra eventualmente escolhida.

Se o modelo de língua não pode dar uma nota a uma tupla de quatro palavras apresentada porque não contém nenhuma combinação dessas palavras, a primeira das quatro palavras será cortada e o resultado de três palavras levado ao modelo, e assim por diante. Portanto, “*to a beautiful woman*” será reduzida a “*a beautiful woman*”, então a “*beautiful woman*”, e finalmente a apenas “*woman*”. Cada um desses quadrigramas, trigramas, bigramas e unigramas será multiplicado por um peso específico para serem classificados. Não é preciso dizer que quanto menor a tupla, menor o peso usado como multiplicador.

A *previsão da próxima palavra* da frase pode representar, do ponto de vista do usuário, uma ajuda muito bem-vinda. Mas, do ponto de vista do programador, é apenas um caso especial da finalização de palavras explicada acima. Para mostrar isso, tomaremos como exemplo a frase “*When I worked on the oil platform I had a terrible*” (Quando eu trabalhei na plataforma de petróleo, eu tive uma terrível). O modelo de língua irá então destacar o contexto “*had a terrible*” (tive uma terrível) e o prefixo “ ”, no qual este consiste de uma linha vazia. Ele irá pegar todas as palavras da lista ordenada mencionada acima, adicionar o contexto e dar a cada uma das tuplas de quatro palavras resultante uma nota. Similar ao exemplo anterior, as adições nas dez tuplas que têm a nota mais alta serão selecionadas e apresentadas ao usuário em uma ordem de frequência. Nesse exemplo em particular, as dez palavras são *time* (tempo), *start* (começo), *experience* (experiência), *impact* (impacto), *effect* (efeito), *accident* (acidente), *season* (temporada), *fight* (luta), *relationship* (relação), e *and* (e). Uma dessas palavras pode satisfazer as necessidades do autor num contexto real. Senão, uma estratégia alternativa tem que ser escolhida, por

exemplo digitar uma palavra em L1. O modelo de língua então garantirá que os equivalentes em L2 oferecidos ao usuário estarão organizados em uma ordem de frequência.

O exemplo seguinte ilustra a *priorização de candidatos à tradução*: Um autor escreveu “*The man was gift*” (O homem era *gift*). A palavra *gift* é polissêmica em dinamarquês e será automaticamente pesquisada em um dicionário dinamarquês-inglês que contém um número de candidatos à tradução, como *poison* (veneno), *venom* (peçonha), *married* (casado), *toxins* (toxinas) etc. Essas palavras são então usadas para substituir *gift* com os resultados seguintes:

“*The man was poison*” (O homem era veneno)

“*The man was venom*” (O homem era peçonha)

“*The man was married*” (O homem era casado)

“*The man was toxins*” (O homem era toxinas)

etc.

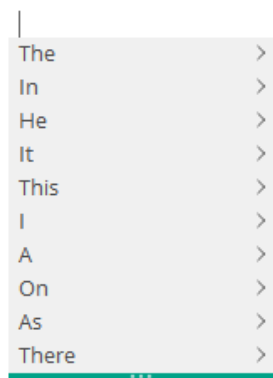
Como foi o caso acima, cada uma dessas tuplas de quatro palavras receberá uma nota do modelo de língua. Felizmente, “*The man was married*” é uma das tuplas de maior nota, o que sugere que *married* é a tradução mais provável para *gift* nesse contexto real. Uma vez que uma nota foi dada, os candidatos à tradução serão apresentados ao usuário na ordem mais provável.

Portanto, o que foi gerado é um *dicionário sensível ao contexto*, i.e. um novo tipo de dicionário que marca um passo adiante em direção a um produto lexicográfico mais personalizado.

4.3 Funcionalidade e exemplos

Quando o usuário abre um novo documento, ou começa um novo parágrafo ou frase, com o Write Assistant funcionando, uma pequena caixa aparecerá com as palavras mais frequentes para iniciar. Se o autor quer usar uma dessas palavras, tudo que ele precisa fazer é clicar nela. Então, a primeira palavra será marcada em verde (vide Figura 1). O autor tem agora duas opções, isto é, usar a palavra ou ir até a palavra certa usando a tecla “para baixo” do teclado. Quando uma das sugestões está marcada em verde, a função normal da tecla “retornar” será desativada e essa tecla pode ser usada para aceitar e inserir a palavra destacada diretamente no texto, então pulando um dos passos no processo de pesquisa de informação tradicional discutido na Seção 1. Assim que a palavra for adicionada, o Write Assistant moverá a janela para frente, para que o prefixo alinhe-se com as primeiras letras das dez sugestões mais prováveis de serem a próxima palavra, e assim por diante (vide Figura 2). Também se beneficiando do modo como o olho funciona.

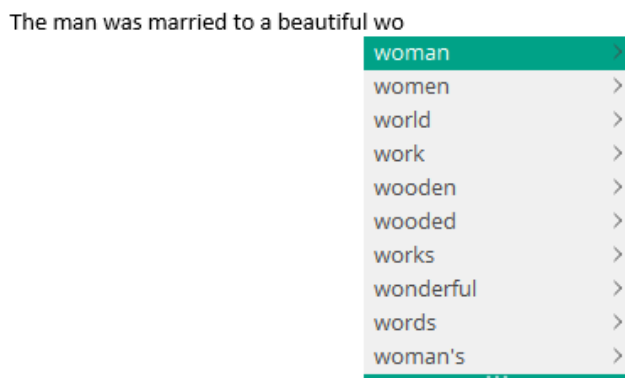
Figura 1 – Palavras iniciais



Fonte: Acervo do autor

Se o autor não optar por nenhuma palavra sugerida e começa a digitar outra, uma caixa com as dez palavras mais prováveis aparecerá, como explicado na Seção 4.2. A Figura 2 mostra o exemplo “*The man was married to a beautiful wo*” usado anteriormente. Entretanto, se uma terceira letra é adicionada e o prefixo se torna “*wom*”, o campo de candidatos possíveis diminuiria e a probabilidade da palavra esperada aparecer entre as de maior nota aumentaria.

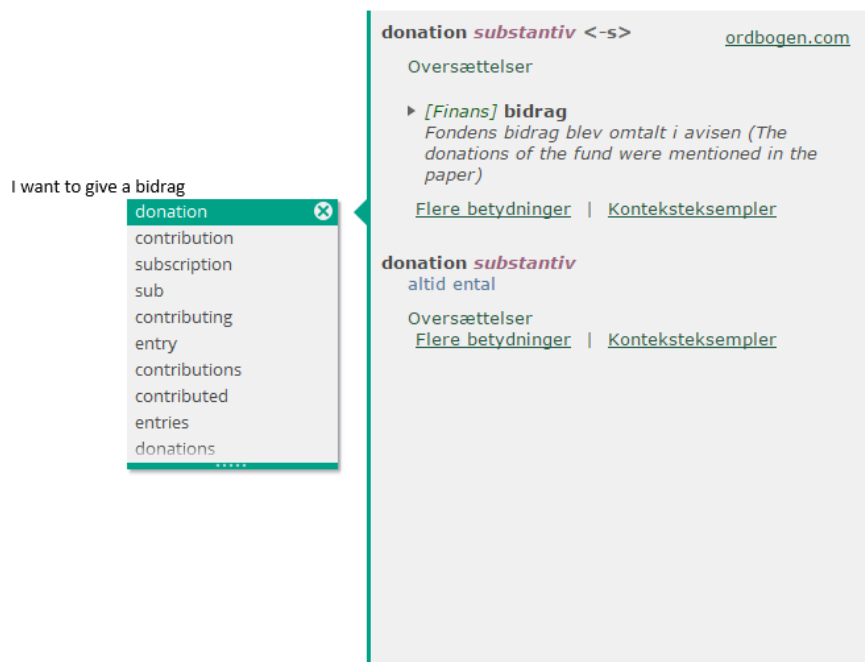
Figura 2 – A tupla “*to be a beautiful wo*” + finalização



Fonte: Acervo do autor

O aplicativo pode também sugerir uma palavra com a qual o autor possa não estar familiarizado, ou nunca tenha visto antes. Nesse caso, a única coisa necessária para conseguir assistência é clicar na seta para a direita da palavra sugerida. O Write Assistant automaticamente pesquisará em um dicionário inglês-dinamarquês e apresentará o resultado em uma nova janela que contém uma entrada de dicionário que mostra definições entre outras informações. Essa funcionalidade pode ser exemplificada pela sequência “*I want to give a bidrag*” (eu quero dar um *bidrag*), na qual *bidrag* é uma palavra polissêmica dinamarquesa com várias equivalentes em inglês, como *donation* (doação), *contribution* (contribuição), *subscription* (inscrição) etc. (vide Figura 3).

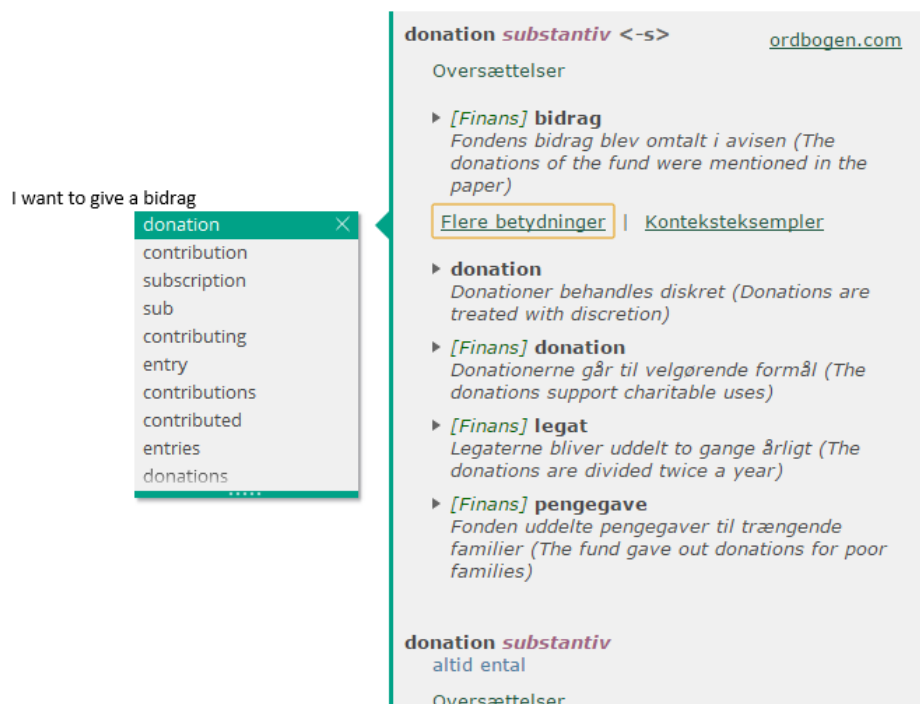
Figura 3 – A tupla “to give a bidrag” + entrada padrão



Fonte: Acervo do autor

A entrada padrão do dicionário reproduzida na Figura 3 fornece apenas um sentido de *donation*. Apesar disso, a janela oferece várias opções ao usuário, que pode clicar em *Flere betydninger* (Mais definições) ou *Konteksteksempler* (Exemplos de contexto) para mais dados (ou até no *ordbogen.com* para ir ao artigo específico do dicionário online). Se o autor optar por *Flere betydninger*, ele receberá acesso instantâneo a mais definições e sentidos como mostrado na Figura 4.

Figura 4 – A tupla “to give a bidrag” + entrada expandida



Fonte: Acervo do autor

Se o problema é que o autor não sabe o significado da palavra sugerida, ou tem dúvidas sobre ela, a necessidade de informação gerada a partir dessa incerteza tem grande probabilidade de ser sanada imediatamente. Entretanto, o problema também pode ser que o autor não sabe como usar a palavra sugerida junto com outras palavras. Nesse caso, uma ajuda adicional pode ser acessada clicando em *Konteksteksempler*. Os dados até agora ocultos no metatexto estarão então visíveis na forma de vários exemplos de sentenças baixados do dicionário inglês-dinamarquês e originalmente selecionados em um *corpus* (vide Figura 5). Isso permitiria ao autor detectar um número de colocações, frases e propriedades sintáticas úteis, entre outras coisas, entretanto esses dados são oferecidos apenas implicitamente. É importante notar que os metatextos *Flere betydninger* e *Konteksteksempler* são fornecidos com dados ocultos como padrão, e que esses dados só podem ser acessados através da decisão do usuário de agir. Dessa maneira, há uma redução do fenômeno crescente e problemático de excesso de informação e dados, por exemplo, a “muralha de texto” que faz as pessoas fugirem instantaneamente de *websites*; vide GOUWS e TARP (2017).

Figura 5 – A tupla “to give a bidrag” + exemplos de sentença

The image shows a screenshot of a translation interface. On the left, a text input field contains the English phrase "I want to give a bidrag". Below it, a dropdown menu displays a list of English equivalents for the Norwegian word "bidrag": "donation", "contribution", "subscription", "sub", "contributing", "entry", "contributions", "contributed", "entries", and "donations". On the right, a larger panel displays the Norwegian word "donation" with its grammatical information "substantiv <-s>" and the source "ordbogen.com". Below this, it lists "Oversættelser" (Translations) with several examples in Norwegian and English. The examples include: "[Finans] bidrag" (Fondens bidrag blev omtalt i avisen / The donations of the fund were mentioned in the paper), "The donation box was forced open at the church...", "Newport Council is now directing anyone wanting to make a donation to the fire victims...", "But, as the firm was registered in the Isle of Man, the donation was impermissible under electoral law.", "The Lyric Theatre's donation from Dr Martin Naughton is thought to be the biggest personal gift made to an arts organisation in NI.", "Labour received £2.33m more than in the previous quarter, following a £2m donation by Britain's richest man, steel magnate Lakshmi Mittal.", and "Hotel owners the Oram family made the".

Fonte: Acervo do autor

Dependendo do nível de proficiência em L2, as pessoas que estão escrevendo textos nessa língua terão que, com alguma frequência, usar sua língua materna para encontrar ou lembrar a palavra em L2 a ser usada. Peguemos o exemplo de um falante nativo de dinamarquês que quer elogiar uma mulher e não sabe, ou não tem certeza, da palavra em inglês que pode expressar essa ideia. Ele então escreve a seguinte sequência: “*She is very smuk*” (Ela é muito *smuk*). *Smuk* é uma palavra polissêmica em dinamarquês com um número relativamente grande de equivalentes em inglês, dos quais vários são mais ou menos sinônimos. O Write Assistant imediatamente procurará no dicionário dinamarquês-inglês e apresentará dez desses equivalentes ao autor em uma ordem priorizada específica (vide Figura 6).

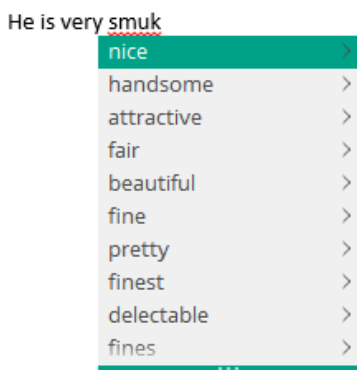
Figura 6 – A tupla “*She is very smuk*” + equivalentes



Fonte: Acervo do autor

Entretanto, se o autor, em vez disso, quisesse expressar algo similar sobre um homem, as palavras digitadas poderiam ser: “*He is very smuk*” (Ele é muito *smuk*). O aplicativo iria então pesquisar no dicionário novamente e sugerir as dez traduções mais prováveis de *smuk* (vide Figura 7).

Figura 7 – A tupla “*He is very smuk*” + equivalentes



Fonte: Acervo do autor

Uma comparação entre as Figuras 6 e 7 mostra outro exemplo de contextualização do dicionário subjacente. Apesar de idênticos nesse caso, os dez equivalentes sugeridos não são apresentados ao usuário na mesma ordem. Alguns desses candidatos à tradução são sinônimos muito próximos e o uso feito pelo autor de um ou de outro faria uma grande diferença. Entretanto, olhemos para as palavras *beautiful* (bela) e *pretty* (linda), essas duas são apresentadas como os equivalentes mais prováveis de *smuk* em relação a *she* (ela), mas aparecem somente na quinta e sétima posição respectivamente em relação a *he* (ele). Em contraste, *nice* (legal) e *handsome* (bonito) são os dois candidatos mais cotados em relação a *he*, enquanto ocupam as posições três e sete respectivamente quando relacionados a *she*. Essa diferença é, sem dúvidas, relevante na medida em que reflete o uso real da língua já

que foi coletada em um *corpus*⁸. De qualquer forma, se o autor ainda não tem certeza sobre qual palavra usar, tudo que ele precisa fazer é clicar na seta para a direita na janela de sugestão para poder consultar explicações sobre as respectivas palavras e, se julgar necessário, clicar nos exemplos de sentença disponíveis no *Konteksteksempler*.

De um ponto de vista filosófico, é interessante observar como a relação entre homem e máquina muda quando o usuário decide clicar na entrada do dicionário ou em um dos metatextos. Até agora, o autor digitou apenas letras e palavras enquanto a ferramenta sugeriu soluções possíveis, as quais o autor pode ignorar ou aceitar com um simples movimento dos dedos. Mais nada lhe foi requisitado. Apesar da interação com a ferramenta, seu usuário basicamente assumiu um papel passivo em termos de assistência fornecida. Porém, quando o usuário clica na seta para a direita, ou clica nos metatextos, ele assume um papel ativo, de quem toma uma decisão consciente de procurar mais informação. Essa situação é muito similar à da consulta tradicional de dicionários independentes, mas com uma diferença importante. Enquanto este tem que ser retirado da estante ou acessado em um *website* separado, os dicionários integrados ao Write Assistant já estão lá, prontos para serem consultados.

4.4 Necessidades de informação parcial ou completamente resolvidas

Nesta seção, tentaremos responder à questão, provavelmente a mais essencial, qual seja, até que ponto o Write Assistant realmente supre as necessidades de informação dos usuários quando eles têm problemas em relação à produção de textos em uma língua estrangeira. Como referência, tomaremos as dez necessidades de informação mais comuns listadas no Quadro 1.

1) Quando um autor pensa que conhece uma palavra em L2, mas não consegue se lembrar dela, ele precisará de uma solução L1-L2 que forneça *traduções em L2* de palavras em L1. Na seção anterior (Figuras 6 e 7), vimos como o aplicativo soluciona o problema completamente, pelo menos para palavras simples. Mas o que acontece com as compostas? Em um assistente para dinamarqueses que estão escrevendo textos em inglês, essa classe de palavra não é um grande problema. A maioria das palavras compostas em dinamarquês são escritas como uma só palavra e, contanto que elas sejam selecionadas como *lemmata* no dicionário L1-L2, os equivalentes em inglês serão automaticamente apresentados aos usuários quando necessários. Entretanto, a maioria das palavras compostas em inglês consiste de duas ou mais palavras com espaços entre elas, uma solução técnica um pouco diferente poderia ser necessária se a ferramenta tivesse sido feita para ajudar falantes nativos de inglês que escrevem em dinamarquês.

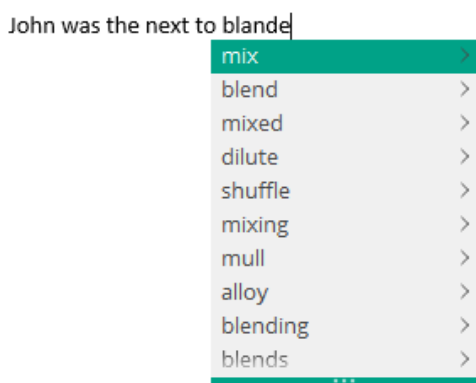
2) Quando um autor não sabe a palavra a ser usada em inglês, ele precisa de uma solução L1-L2 que forneça a tradução em L2 assim como *diferenciação de significado*. Esse

⁸ N.T.: *Beautiful*, *pretty* e *handsome* são palavras usadas para falar sobre a beleza de alguém. *Handsome* é, geralmente, associada a pessoas do sexo masculino, enquanto *beautiful* e *pretty* são associadas a pessoas do sexo feminino.

problema também é completamente solucionado pelo Write Assistant, mas de uma maneira que difere da solução tradicional nos dicionários, nos quais os diferenciadores de significado, via de regra, são visíveis simultaneamente. Neste aplicativo, os dez equivalentes mais prováveis são fornecidos em uma ordem priorizada (uma grande vantagem em comparação aos dicionários tradicionais), mas a diferenciação deve ser acessada separadamente para cada equivalente. Nesse caso, as *explicações* das respectivas palavras em L2 (agora apresentadas como equivalentes) servem como diferenciadores de significado (vide Figura 4).

3) Quando um autor não conhece um colocação em L2 a ser usada, ele precisará de uma solução L1-L2 que forneça *traduções em L2* de colocações em L1. Em sua versão atual, o Write Assistant não oferece essa solução. Ele funciona relativamente bem quando as colocações são diretas com bases e colocados semelhantes, de acordo com a teoria de colocação de duas palavras de Hausmann (1985). Se o autor dinamarquês, por exemplo, digitar o verbo *børste*, o aplicativo irá sugerir *brush* (escovar) como o primeiro candidato à tradução, e se o autor continuar com *mine tænder* (meus dentes), não haveria grande problema. Mas se a L2 for espanhol, onde a colocação equivalente é *lavarme los dientes* (lavar-me os dentes, literalmente), não haveria solução direta. Mas mesmo que haja uma colocação em L1 mais ou menos parecida, não significa necessariamente que a solução seja fácil. Tomemos a colocação dinamarquesa *blande kort* (embaralhar as cartas) como exemplo. Quando o autor digita a palavra altamente polissêmica *blande*, por exemplo na sequência “*John was the next to blande*” (João foi o próximo a blande), várias traduções de *blande* serão propostas pela ferramenta. No entanto, *shuffle* (embaralhar) só aparecerá como o quinto candidato mais provável, depois de *mix* (misturar), *blend* (combinar), *mixed* (misturado) e *dilute* (diluir) (vide Figura 8). O risco de escolher o equivalente errado em inglês provavelmente seria alto se o autor não estivesse muito consciente sobre o problema e buscasse assistência adicional no dicionário integrado, onde *shuffle* é indicado como o verbo correto em conexão com *cards* (cartas).

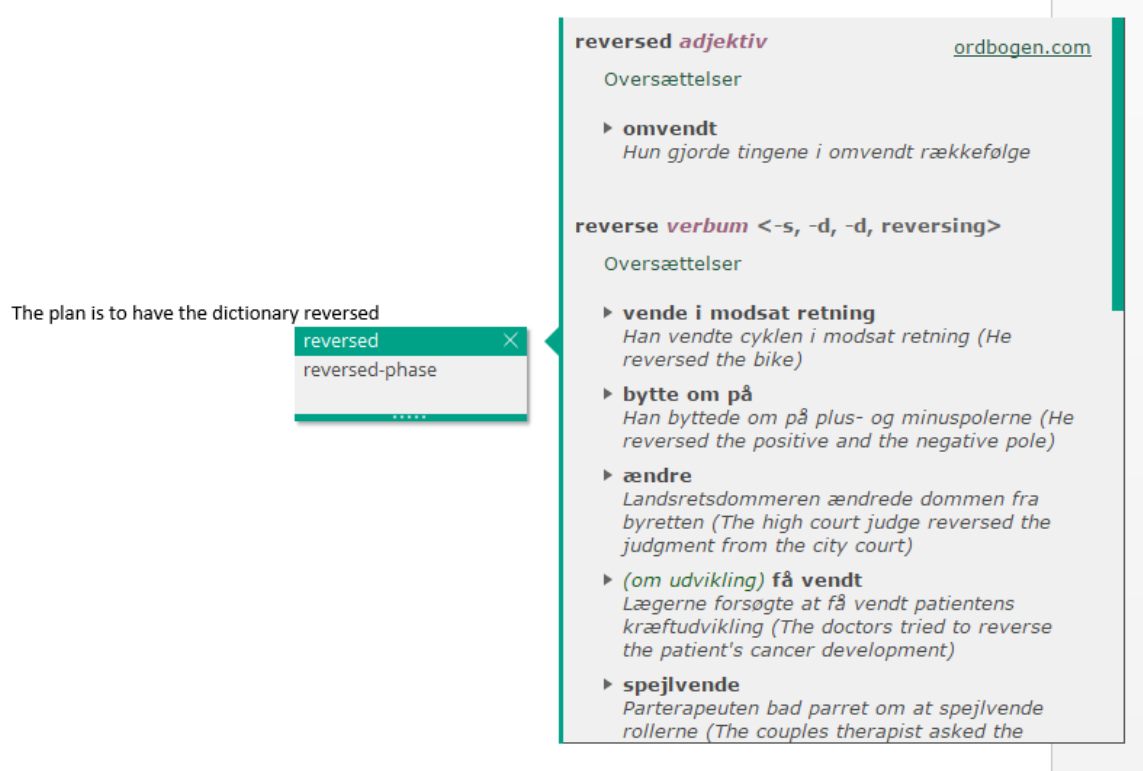
Figura 8 – A tupla “the next to blande” + equivalentes



Fonte: Acervo do autor

4) Quando um autor conhece uma palavra em L2, mas não tem certeza se ela pode ser usada com o significado concreto, ele precisará de *explicações* das palavras em L2. Esse problema é completamente solucionado, como pode ser visto na Figura 9.

Figura 9 – A tupla “have the dictionary reversed” + entrada do dicionário



Fonte: Acervo do autor

5) Quando o autor conhece uma palavra em L2, mas não tem certeza se ela pode ser usada em um contexto real, ele precisará de *informações estilísticas, pragmáticas, culturais* ou *de políticas linguísticas*. Esse problema é solucionado apenas em um grau limitado. O dicionário L2-L1 integrado na ferramenta não oferece dados explícitos desse tipo. Mas a informação pode, até certo ponto, ser deduzida dos exemplos de sentenças disponíveis no *Konteksteksempler*. Como um caso especial, o Write Assistant também estará disponível em versões especificamente adaptadas para empresas reais que tenham suas próprias políticas linguísticas em termos da terminologia a ser usada por seus funcionários. Essas versões especiais do aplicativo serão, portanto, projetadas para refletir e transmitir as informações relevantes de políticas linguísticas.

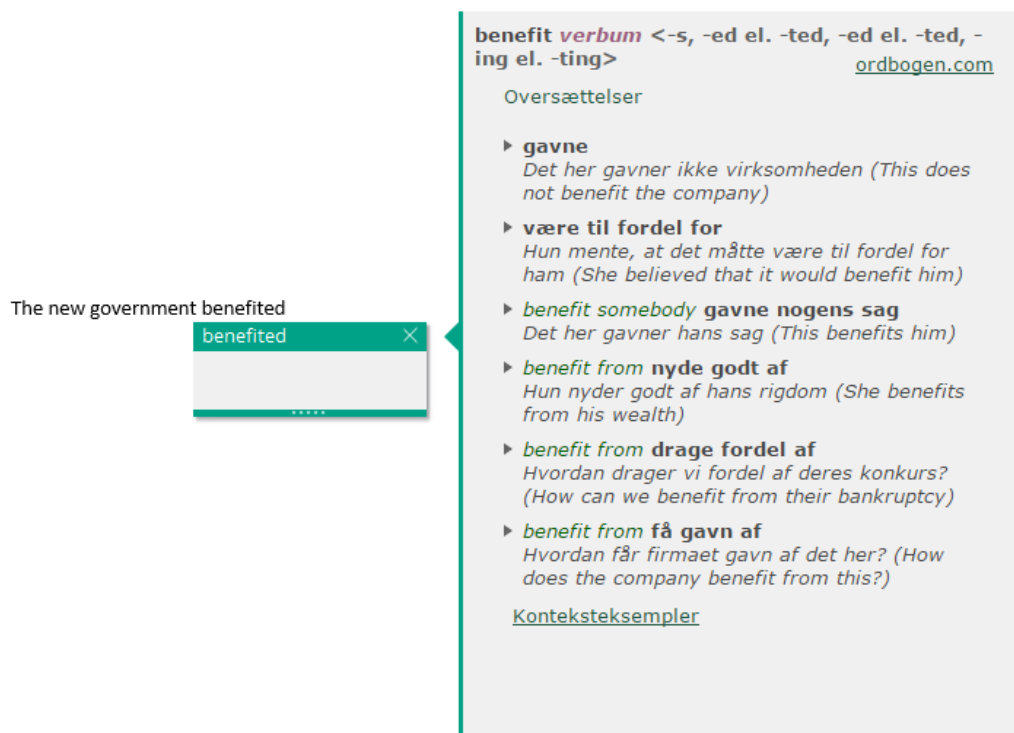
6) Quando o autor conhece uma palavra em L2, mas não tem certeza de como ela é escrita, ele precisará de informações sobre *ortografia* ou *autocorreção* ao digitar palavras em L2. Esse problema é completamente solucionado. Por um lado, a ferramenta oferece finalizações escritas corretamente depois que as primeiras letras são digitadas e, por outro lado, funciona junto com o verificador de ortografia e gramática da Microsoft, que detecta possíveis erros de ortografia e sugere correções. Alternativamente, o usuário poderia

simplesmente escrever uma palavra em L1 para obter a palavra em L2 escrita corretamente.

7) Quando o autor conhece uma palavra em L2, mas não tem certeza de como ela é flexionada, ele precisará de informações sobre *formas flexionadas*. Esse problema é completamente solucionado. As possíveis finalizações, próximas palavras e equivalentes sugeridas pelo aplicativo são todas flexionadas para se encaixar na sentença, mas apenas por estatísticas no modelo de linguagem, isso é, sem aplicar regras gramaticais. Se informações adicionais forem necessárias, a tabela de flexão completa da palavra em questão pode ser acessada no dicionário L2-L1 (vide Figura 9). De qualquer forma, o verificador de ortografia e gramática da Microsoft também estará presente como uma garantia medianamente eficaz.

8) Quando o autor conhece uma palavra em L2, mas não sabe como combiná-la com outras palavras, ele precisará de informações explícitas e/ou implícitas sobre as *propriedades sintáticas* dela. Essa é uma das questões mais complexas da Lexicografia. Como quase todos os dicionários existentes, o Write Assistant não apresenta uma resposta completamente satisfatória, mas fornece pelo menos alguma ajuda. Como pode ser visto na Figura 10, o dicionário L2-L1 oferece algumas “regras” sintáticas que são posteriormente exemplificadas. Alguma assistência adicional também pode ser encontrada nos exemplos (na Figura 5), mas os usuários têm que deduzir as regras subjacentes quando querem construir suas próprias sentenças. No entanto, nem os dados explícitos nem os implícitos são direcionados a sentidos específicos, um ponto fraco que deixa espaço para erros. O que é necessário para satisfazer as necessidades de diferentes tipos de usuários é uma combinação de dados explícitos (regras) e implícitos (exemplos), todos direcionados aos sentidos específicos da palavra. Esse requisito é atendido apenas parcialmente pelo aplicativo em sua versão atual.

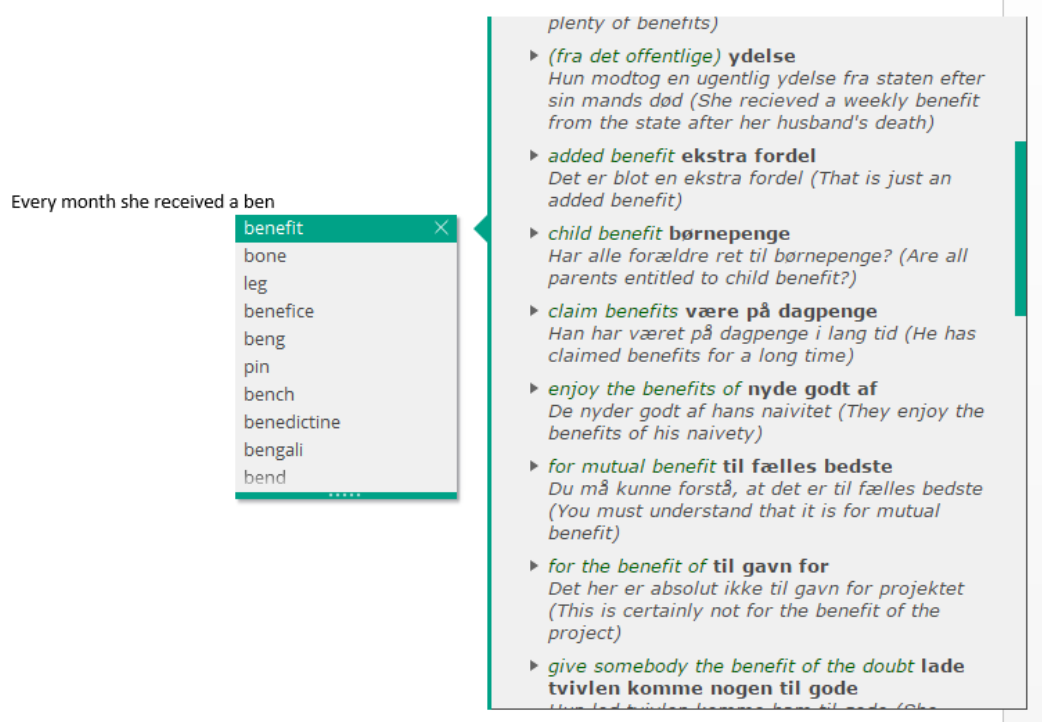
Figura 10 – A tupla “*The new government benefited*” + entrada do dicionário



Fonte: Acervo do autor

9) Quando o autor conhece uma palavra em L2, mas não sabe como construir colocações com essa palavra, ele precisará de *colocações* com a palavra em L2 em questão. Esse problema se sobrepõe ao discutido no ponto 3. A principal diferença é que os autores, neste caso, conhecem pelo menos uma das palavras que compõem a colocação. O problema poderia, portanto, ser resolvido consultando um dicionário baseado em L2 que contenha colocações. Como pode ser visto na Figura 11, o dicionário integrado na ferramenta oferece algumas colocações explícitas, além daquelas que podem ser deduzidas dos exemplos. Além disso, se os autores conhecem a primeira palavra da colocação e a digitam, é possível que a segunda palavra esteja entre aquelas sugeridas pela ferramenta, pelo menos se for uma colocação frequentemente usada. Se não for assim, ou se for apenas a segunda palavra que é conhecida pelo escritor, então é necessária uma solução L1-L2 como no exemplo com *shuffle the cards* discutida acima. Tal solução atualmente não é fornecida. Por isso, o Write Assistant ainda não apresentou uma resposta completamente satisfatória aos desafios colocados para uso de colocações em L2.

Figura 11 – A tupla “*she received a ben*” + entrada do dicionário



Fonte: Acervo do autor

10) Quando o autor conhece uma palavra em L2, mas quer variar o vocabulário e usar outra palavra, ele precisará de *sinônimos* e/ou *antônimos*. Esse problema não é completamente atendido. Quando a ferramenta sugere possibilidades de próximas palavras a serem escritas na frase, às vezes há sinônimos entre as palavras propostas, mas essa solução pode, por definição, nunca ser sistemática, na medida em que as palavras apresentadas ao usuário pelo modelo de língua não dependem de suas estruturas semânticas, mas da probabilidade das respectivas tuplas serem encontradas em L2. Assim, uma solução mais apropriada seria que o dicionário subjacente oferecesse sinônimos e antônimos, um serviço que ele poderia facilmente realizar.

Em suma, pode-se concluir que a primeira versão do Write Assistant atende completamente à metade dos dez requisitos enquanto ferramenta projetada para auxiliar na escrita em L2, mas apenas parcialmente à outra metade. As limitações podem ser devidas à tecnologia desenvolvida, ao design da interface ou à qualidade da base empírica. Entre os cinco requisitos que não encontraram solução satisfatória, apenas aqueles relacionados às colocações em L2 podem ser explicados por limitações tecnológicas. O desafio é, portanto, desenvolver ainda mais o Write Assistant na medida em que ele não apenas complete ou preveja palavras com base no contexto das três palavras anteriores, mas também analise o texto anterior e proponha correções de palavras já digitadas para oferecer as colocações corretas. Tal solução também seria relevante para compostos e termos de múltiplas palavras e, simultaneamente, aumentaria a consciência de contexto do dicionário integrado.

Os problemas remanescentes detectados acima – ou seja, aqueles relacionados ao tratamento de sinônimos, antônimos, sintaxe, estilo, pragmática e cultura – exigem uma solução diferente na medida em que têm a ver com o design das entradas de dicionário visualizadas e com o conteúdo do dicionário subjacente usado como recurso empírico. Esses problemas são todos de natureza lexicográfica e serão brevemente discutidos na Seção 5.

4.5 Utilidade e fluxo da escrita

Na introdução deste artigo nos referimos ao processo tradicional de busca de informações, como descrito por Bergenholtz *et al.* (2015). Ressaltamos a necessidade de uma ferramenta que auxilie na escrita em L2 de modo que algumas das fases e etapas desse processo possam ser encurtadas ou mesmo ignoradas, a fim de economizar tempo e manter o fluxo de escrita sem perder o foco na mensagem a ser transmitida. A análise que subsequentemente fizemos do Write Assistant indica que ele é um candidato qualificado para se tornar uma ferramenta com essas propriedades. Como se pode esperar de versões beta de produtos digitais complexos, ele possui alguns problemas de nascença, mas todos parecem ser curáveis.

A ferramenta parece ser muito simples e fácil de usar do ponto de vista funcional. Seus usuários podem trabalhar com seu teclado normal, navegar entre as palavras sugeridas e realizar pesquisas complementares nos dicionários integrados, usando o mouse ou a tecla *Alt* e as setas. Ademais, a tecla de retorno foi reprogramada para que possa ser usada para inserir palavras diretamente no texto quando estão marcadas em verde. Tudo isso cria as condições técnicas para um fluxo de escrita aprimorado.

Os autores podem se tornar mais conscientes de suas necessidades de informação quando começam a digitar e o aplicativo imediatamente apresenta sugestões para finalizações e próximas palavras. Se eles estiverem satisfeitos com uma dessas sugestões, a única ação necessária é marcar a palavra em questão e pressionar a tecla de retorno. O mesmo acontece quando eles digitam uma palavra em L1 e instantaneamente obtêm um número de candidatos em L2. Se mais informação for necessária para quaisquer das palavras sugeridas, basta um simples movimento do dedo para imediatamente visualizar um texto sobre a palavra no dicionário. Uma vez que o resultado seja considerado satisfatório e a tecla de retorno for pressionada, a palavra em questão será automaticamente adicionada ao texto e resolverá o problema que originalmente criou a necessidade de informação.

Como pode ser visto no que foi apresentado acima, algumas das fases e etapas tradicionais do processo de busca de informações tornaram-se menos complexas, enquanto outras foram ignoradas porque a necessidade de acessar recursos de informações externas foi consideravelmente reduzida. Na verdade, os usuários do aplicativo podem percorrer um longo caminho sem consultar esses recursos. Eles podem economizar um tempo precioso e se concentrar na mensagem a ser escrita.

Por outro lado, enquanto o Write Assistant facilita, sem dúvidas, o trabalho dos autores, ele não retira nenhuma de suas responsabilidades. Os usuários desta ferramenta ainda devem avaliar criticamente as informações recuperadas, decidir quais palavras usar e assumir a responsabilidade pelo texto final. Os usuários, e mais ninguém, são os únicos autores dos textos produzidos com o auxílio da ferramenta. O aplicativo é apenas um assistente útil, não um coautor.

5. Desafios para a lexicografia

A existência de computadores e grandes bancos de dados, a programação do modelo de língua e o design funcional e amigável compõem as condições tecnológicas e técnicas necessárias para um assistente de escrita em L2 como o descrito aqui. No entanto, essas condições não garantem por si só a qualidade do aplicativo. Uma vez criadas, a qualidade do produto depende, antes de tudo, da quantidade e da qualidade da base empírica a partir da qual os dados são recolhidos, isto é, o *corpus* e os dicionários.

A exigência para o *corpus* é basicamente que ele deva ser relativamente grande, bem composto, atualizado e suficientemente “limpo” para reduzir o risco de erros ortográficos. Mas e os dicionários? Muitos dos problemas detectados na Seção 4.4 têm a ver com os dicionários subjacentes. Uma análise mais detalhada dos dados lexicográficos confirmaria essa tendência. A crítica da quantidade de palavras em L1 com os equivalentes em L2, a quantidade e qualidade dos equivalentes, a qualidade das definições das palavras em L2, a existência ou não de outras categorias de dados relevantes, bem como a qualidade dessas - tudo isso é, em grande medida, uma crítica aos dicionários utilizados como recursos empíricos.

Não é difícil encontrar (ou compilar) dicionários que contenham a maioria das categorias de dados faltantes que foram discutidas na Seção 4.4. Com a possível exceção de notas culturais apropriadas, os outros itens – ou seja, rótulos estilísticos e pragmáticos, sinônimos e antônimos, colocações assim como dados sintáticos explícitos e implícitos – já estão disponíveis em muitos dicionários existentes, pelo menos em uma língua como o inglês. No entanto, todos esses itens devem ser adaptados aos requisitos técnicos e funcionais específicos da ferramenta, uma adaptação que pode ter consequências para o conceito geral do dicionário e para os métodos de produção em novos projetos de dicionário. Tudo isso impõe novos desafios à Lexicografia.

O uso de dicionários já existentes pode criar alguns problemas que afetam a qualidade geral da ferramenta. Hoje, a maioria dos dicionários ainda é concebida para fornecer assistência à produção de textos e à recepção de textos sem a necessária diferenciação. A concepção e a apresentação dos diferentes itens lexicográficos utilizados para ambos os fins podem, portanto, representar uma espécie de acordo que não é necessariamente o mais adequado para a escrita em L2. Além disso, pode haver outros problemas em termos de escopo e direcionalidade. O dicionário utilizado para sustentar o Write Assistant é *bidirecional*, isto é, consiste num conjunto de dicionários bilíngues nas direções de ambas as línguas com o requisito especial de que todas as palavras em L2 no

dicionário L1-L2 devem ter uma entrada equivalente no dicionário L2-L1. A solução bidirecional parece ser a mais apropriada em conexão com a escrita em L2, e é também a recomendada para esse propósito pela maioria dos lexicógrafos modernos; vide, por exemplo, LEW e ADAMSKA-SALACIAK (2015). Mesmo assim, é importante que o dicionário seja projetado do zero como um dicionário *monodirecional*, ou seja, para usuários de apenas uma das duas línguas. A razão para esse requisito conceitual fundamental é que cada versão do Write Assistant possui apenas um grupo de usuários em termos de língua materna, ou seja, falantes nativos de L1, e que qualquer bidirecionalidade pode interferir de maneira inconveniente com os dados lexicográficos apresentados a esse grupo específico de usuários.

Aqui, discutiremos brevemente os *itens de significado* como apenas um exemplo de como os dados lexicográficos obtidos de um conjunto de dicionários bilíngues não totalmente adaptados podem não ser a melhor solução. Nos dicionários bilíngues tradicionais existem duas classes principais de itens de significado em termos de propósito: 1) o que é usado para diferenciar entre os equivalentes de L2 no dicionário L1-L2, e 2) o que é usado para explicar o significado de uma palavra em L2 no dicionário L2-L1. Diferentes técnicas são aplicadas em cada caso. Nos dicionários L2-L1, nos quais o foco está na recepção do texto, os itens mais frequentemente usados para explicar o significado das palavras em L2 são equivalentes em L1, enquanto os itens usados para diferenciar entre os equivalentes em L2 nos dicionários L1-L2 são tipicamente sinônimos e paráfrases em L1.

Na Seção 4.4, vimos como a mesma categoria de dados desempenhou o papel de ambos, explicação de significado e diferenciador de significado. Esse duplo papel pode criar alguns inconvenientes se não tiver sido previsto e levado em consideração no momento em que o conceito para os dicionários subjacentes foi decidido. Figuras 12 e 13 ilustram o problema.

Figura 12 – A tupla “Internet has some *fordele*” + entrada do dicionário

Compared to corpora, the use of internet has some *fordele*

- advantages
- advantage
- shared
- shares
- share
- spread
- sharing
- distributed
- dispersed
- spreading

advantage *substantiv* <-s> ordbogen.com
Oversættelser
► **fordel**
Vi har opnået en afgørende fordel
[Flere betydninger](#) | [Konteksteksempler](#)
advantage *verbum* <-s, -d, -d, advantaging>
Oversættelser
[Flere betydninger](#)

Fonte: Acervo do autor

Figura 13 – A tupla “Internet has some *ad*” + entrada do dicionário

Compared to corpora, the use of Internet has some *ad*

- advantages
- additional
- advice
- advantage
- advanced
- additions
- by
- at
- added
- admins

advantage *substantiv* <-s> ordbogen.com
Oversættelser
► **fordel**
Vi har opnået en afgørende fordel
► **overtag**
Han forstod at bruge sit overtag
► **fortrin**
Den nye kollektions fortrin er iøjensfaldende
► **forspring**
Hun udnyttede sit forspring på bedste vis
► **have the advantage of have en fordel frem for**
Hendes rige baggrund gør, at hun har en fordel frem for konkurrenten
► **take advantage of udnytte**
Sælgeren udnyttede de godtroende kunder
► **take advantage of benytte sig af**
Hun vægrede sig ved at benytte sig af chefens tilbud
► **take advantage of drag fordel af**
De forstod at drage fordel af den nuværende lovgivning
► **to somebody's advantage til nogens fordel**
Handlen var helt klar til hans fordel

Fonte: Acervo do autor

Embora retirados do mesmo dicionário subjacente, as duas entradas padrão

reproduzidas nas Figuras 12 e 13 são diferentes porque apenas um item de significado é necessário como um *diferenciador* no primeiro, enquanto vários itens são necessários para *explicar* o significado de *advantages* (vantagens) no segundo. Em ambos os casos, os itens de significado consistem de uma palavra em L1 exemplificada por uma frase em L1 (por exemplo, *fordel* – *Vi har opnået en afgørende fordel*). Esse tipo de item pode ser útil quando autores têm dúvidas sobre o significado de uma palavra em L2. Mas eles provavelmente ficarão mais confusos do que esclarecidos quando uma palavra em L1 (neste caso *fordel*) for usada para diferenciar entre seus próprios equivalentes, embora o exemplo seguinte possa solucionar o problema até certo ponto. Uma solução diferente é, portanto, necessária, de preferência uma definição curta em L1 que seria útil tanto como explicação quanto como diferenciador.

Itens de significado não são exceção. Outras classes de dados lexicográficos também precisam ser escrutinadas e adaptadas aos requisitos especiais do Write Assistant e ferramentas similares como uma pré-condição para otimizar a qualidade do serviço prestado.

6. Perspectivas

No primeiro semestre de 2017, o Write Assistant foi testado entre um pequeno grupo de estudantes dinamarqueses do ensino médio, e também foi demonstrado para professores de inglês e estudantes de várias universidades chinesas. O *feedback* foi geralmente muito positivo em termos de utilidade global, mas foram expressas opiniões diferentes sobre as possíveis consequências para a aprendizagem de língua estrangeira. Por exemplo, expressou-se o medo de que futuros estudantes de idiomas possam se tornar muito dependentes da ferramenta. Isso pode ser verdade. Um medo similar foi expresso quando a calculadora foi introduzida no ensino de matemática. Hoje, é óbvio que muitas pessoas são altamente dependentes da calculadora, mas também é um fato que ela lhes permite realizar cálculos mais complexos do que nunca sem cometer muitos erros.

O uso do Write Assistant e de ferramentas similares requer a consciência do papel das pessoas e da máquina na comunicação moderna. Ainda é a pessoa que é a única responsável pelo conteúdo e pela forma da mensagem a ser escrita. A tecnologia está aqui apenas para ajudar, não para assumir o controle. Compreendendo isso, ferramentas como o Write Assistant definitivamente serão úteis na escrita em L2. Por outro lado, estudantes que atualmente continuam usando a tradução automática de forma não crítica, apesar de serem repetidamente avisados por seus professores, provavelmente não serão os que mais se beneficiem dessas ferramentas.

Não há dúvidas de que a nova tecnologia veio para ficar. Ferramentas de alta tecnologia projetadas para auxiliar a escrita, leitura e tradução de textos serão parte integrante de nossas vidas nos próximos anos. As pessoas ficarão cada vez mais dependentes delas, gostemos ou não. A Lexicografia pode se adaptar a essa realidade ou morrer. Os lexicógrafos são, portanto, desafiados, não apenas para levantar novas questões

e possibilidades, mas também para considerar questões antigas a partir de um novo ângulo. Não há perspectiva em transformar a disciplina em um Cavaleiro da Triste Figura.

Referências

BERGENHOLTZ, H.; BOTHMA, T.D.J. & GOUWS, R.H. Phases and Steps in the Access to Data in Information Tools. *Lexikos*, v. 25, p. 1-30, 2015.

BOGAARDS, Paul. Dictionaries and productive tasks in a foreign language. *Kernerman Dictionary News*, v. 13, p. 20-23, 2005.

BOWKER, L. Meeting the Needs of Translators in the Age of e-Lexicography: Exploring the Possibilities. In: GRANGER, S.; PAQUOT, M. (Org.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012. p. 379–397.

CHON, Y. V. The Electronic Dictionary for Writing: A Solution or a Problem? *International Journal of Lexicography*, v. 22, n. 1, p. 23–54, 2009.

EINSTEIN, A.; INFELD, L. *The evolution of physics*. New York: Simon & Schuster, 1938.

FUERTES-OLIVEIRA, P. A. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. *International Journal of Lexicography*, v. 29, n. 2, p. 226–247, 2016.

FUERTES-OLIVEIRA, P. A.; TARP, S. *Theory and practice of specialised online dictionaries: Lexicography versus terminography*. Berlin, Boston: De Gruyter, 2014.

GOUWS, R. H.; TARP, S. Information overload and data overload in lexicography. *International Journal of Lexicography*, v. 30, n. 4, p. 389–415, 1 dez. 2017.

GRANGER, S.; PAQUOT, M. Electronic Lexicography Goes Local: Design and Structures of a Needs-driven Online Academic Writing Aid. *Lexicographica*, v. 31, n. 1, p. 118–141, 2015.

HANKS, P. Lexicography, Printing Technology, and the Spread of Renaissance Culture. In: DYKSTRA, A.; SCHOONHEIM, T. (Org.). *Proceedings of the XIV Euralex International Congress*. Leeuwarden: Fryske Akademy, 2010. p. 988–1016.

HAUSMANN, F. J. Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: BERGENHOLTZ, H.; MUGDAN, J. (Org.). *Lexikographie und Grammatik*. Tübingen: Niemeyer, 1985. p. 118–129.

JOHNSON, S. *The Plan of an English Dictionary*. Disponível em: <<https://andromeda.rutgers.edu/~jlynch/Texts/plan.html>>. Acesso em: 31 maio 2017.

LEW, R. Can a Dictionary Help You Write Better? A User Study of an Active Bilingual Dictionary for Polish Learners of English. *International Journal of Lexicography*, v. 29, n. 3, p. 353–366, 2016.

LEW, R.; ADAMSKA-SALACIAK, A. A Case for Bilingual Learners' Dictionaries. *ELT Journal*, v. 69, n. 1, p. 47–57, 2015.

NESI, H. The Demands of Users and the Publishing World: Printed or Online? Free or Paid for? In: DURKIN, P. (Org.). *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, 2015. p. 579–589.

NIELSEN, S. Lexicography and Interdisciplinarity. In: FUERTES-OLIVEIRA, P. A. (Org.). *Routledge Handbook of Lexicography*. London: Routledge, 2017. p. 93–104.

NIELSEN, S. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos*, v. 18, p. 170–189, 2008.

NOMDEDEU-RULL, A.; TARP, S. Hacia un modelo de diccionario en línea para aprendices de español como LE/L2. *Journal of Spanish Language Teaching*, v. 5 (1), n. 9, p. 50–65, 2018.

PAQUOT, M. The LEAD Dictionary-cum-writing Aid: an Integrated Dictionary and Corpus Tool. In: GRANGER, S.; PAQUOT, M. (Org.). *Electronic lexicography*. Oxford: Oxford University Press, 2012. p. 163–185.

ROBINSON, H.; WYSOCKA, A.; HAND, C. Internet Advertising Effectiveness. The Effect of Design on Click-Through Rates for Banner Ads. *International Journal of Advertising*, v. 26, n. 4, p. 527–541, 2007.

RUNDELL, M. Dictionary Use in Production. *International Journal of Lexicography*, v. 12, n. 1, p. 35–53, 1999.

_____. The Dictionary of the Future. In: GRANGER, S. (Org.). *Optimizing the Role of Language in Technology-enhanced Learning*. [S.l.]: NOE-Kaleidoscope, 2007. p. 49–51. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00197203/document/>>. Acesso em: 31 maio 2017.

RUNDELL, M. What Future for the Learner's Dictionary? In: KERNERMANN, I.; BOGAARDS, P. (Org.). *English Learners' Dictionaries at the DSNA 2009*. Jerusalem: Kdictionaries, 2010. p. 169–175.

RUNDELL, M.; KILGARRIFF, A. Automating the Creation of Dictionaries: Where Will It All End? In: MEUNIER, F. et al. (Org.). *A Taste for Corpora. In Honour of Sylviane Granger*. Amsterdam, Philadelphia: John Benjamins, 2011. p. 257–282.

TARP, S. Beyond Lexicography: New Visions and Challenges in the Information Age. In: BERGENHOLTZ, H.; NIELSEN, S.; TARP, S. (Org.). *Lexicography at a Crossroads: Dictionaries and Encyclopedias today, Lexicographical Tools tomorrow*. Bern: Peter Lang, 2009. p. 17–32.

_____. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In: FUERTES-OLIVEIRA, P. A.;

BERGENHOLTZ, H. (Org.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London, New York: Continuum, 2011. p. 54–70.

_____. Lexicography as an Independent Science. In: FUERTES-OLIVEIRA, P. A. (Org.). *Routledge Handbook of Lexicography*. London: Routledge, 2017. p. 19–33.

_____. *Lexicography in the Borderland between Knowledge and Non-knowledge*. Tübingen: Niemeyer, 2008.

_____. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos*, v. 14, p. 299–325, 2004.

TARP, S.; FISKE, K.; SEPSTRUP, P. L2 write assistant and context-aware dictionaries: New challenges to lexicography. *Lexikos*, v. 27, p. 494–521, 2017.

VERLINDE, S. Modelling Interactive Reading, Translation and Writing Assistants. In: FUERTES-OLIVEIRA, P. A.; BERGENHOLTZ, H. (Org.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London, New York: Continuum, 2011. p. 275–286.

VERLINDE, S.; LEROYER, P.; BINON, J. Search and You Will Find. From Stand-Alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. *International Journal of Lexicography*, v. 23, n. 1, p. 2–17, 2010.

VERLINDE, S.; PEETERS, G. Data Access Revisited: the Interactive Language Toolbox. In: GRANGER, S.; PAQUOT, M. (Org.). *Electronic lexicography*. Oxford: Oxford University Press, 2012. p. 147–162.

WANNER, L.; VERLINDE, S.; ALONSO RAMOS, M. Writing Assistants and Automatic Lexical Error Correction: Word Combinatorics. In: KOSEM, I. et al. (Org.). *Electronic Lexicography in the 21st Century: Thinking Outside the Paper*. Ljubljana: Institute for Applied Slovene Studies, 2013. p. 472–487.

Como citar este texto (ABNT):

TARP, S.; FISKE, K.; SEPSTRUP, P. Tradução de Guilherme S. de Oliveira. Dicionários sensíveis ao contexto e integrados a assistentes de escrita em L2: novos desafios para a lexicografia. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 33-62, 2018.

DICIONÁRIOS E EXEMPLOS DE APOIO À PRODUÇÃO LINGUÍSTICA¹

Ana Frankenberg-Garcia²

Tradução: Luísa Rabaldo³

Revisão: Rozane R. Rebechi⁴

Resumo: Os dicionários de aprendizes não só ensinam os significados das palavras, como também incluem informações sobre colocações, sintaxe e erros comuns. Embora existam várias maneiras de se apresentarem informações de apoio à produção linguística nos dicionários, pesquisas mostram que os alunos preferem aprender a partir de exemplos. No entanto, as primeiras pesquisas sobre os efeitos dos exemplos em dicionários não produziram resultados significativos. Uma explicação possível seria porque os dicionários não costumam distinguir entre exemplos que reforçam a compreensão e exemplos que auxiliam na produção linguística. Outra hipótese seria que os alunos podem precisar de mais do que um exemplo para consolidar a compreensão. Dando continuidade aos recentes estudos de Frankenberg-Garcia (2012, 2014), que mostram que exemplos específicos de apoio à produção linguística parecem ter um efeito positivo, o presente trabalho tem como objetivo investigar mais a fundo os efeitos desse tipo de exemplo. Nesse sentido, procurou-se aprimorar os procedimentos de elicitação anteriores, replicar o estudo com uma amostra populacional nova e maior e avaliar se os alunos reagem melhor perante um número ideal de exemplos.

1. Introdução

Tradicionalmente, os dicionários são usados para consultar o significado ou a ortografia das palavras. No entanto, hoje em dia muitos dicionários fornecem informações adicionais, que vão além de grafias e definições (ou traduções, no caso de dicionários bilíngues). Nesse sentido, a lexicografia pedagógica da língua inglesa é particularmente avançada. Dicionários de última geração apresentam dados linguísticos compilados por

¹ A presente tradução foi autorizada para ser publicada em português pela autora Ana Frankenberg-Garcia. Referência bibliográfica completa do artigo original: FRANKENBERG-GARCIA, A. Dictionaries and encoding examples to support language production. In: *International Journal of Lexicography*, 28(4), 2015, p. 490-512.

² Centro de Estudos de Tradução, Faculdade de Inglês e Línguas, Universidade de Surrey (a.frankenberg-garcia@surrey.ac.uk).

³ Bacharelada em Letras – Português/Inglês e bolsista de Iniciação Científica, UFRGS.

⁴ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

meio de *corpora*, com informações sobre colocações e sintaxe, e até sobre como evitar erros comuns.

1.1 Informações de codificação em dicionários

Existem diversas formas de se apresentarem informações de apoio à produção linguística nos dicionários. Como mostra a Tabela 1, isso inclui detalhes gramaticais exibidos de diferentes maneiras. Também se observa que é normal combinar mais de um método de apresentação. Apesar desses avanços em lexicografia, o usuário comum não sabe utilizar tudo que os dicionários oferecem, deixando de aproveitar sobretudo o tipo de informação gramatical, em foco na Tabela 1¹. Um experimento envolvendo 211 alunos de graduação de uma universidade portuguesa, realizado por Frankenberg-Garcia (2011), mostrou que apenas uma pequena minoria recorreria a um dicionário para descobrir qual preposição usar com determinado verbo intransitivo. Em um estudo anterior, Béjoint (1981) aplicou um questionário para investigar o uso de dicionários por um grupo de 122 alunos franceses de inglês, e concluiu que menos da metade dos entrevistados consultava informações sobre a sintaxe das palavras. Mesmo os alunos de graduação em Letras, que supostamente estariam mais atentos a questões sintáticas, não parecem se sair melhor. Herbst e Stein (1987) relataram que muitos estudantes universitários alemães de língua inglesa, e até mesmo alguns professores, não sabiam que os dicionários continham informações sobre os padrões sintáticos das palavras.

Em outra investigação sobre a utilização de dicionários na escrita, Dziemianko (2006) conduziu um estudo em larga escala com 606 alunos do ensino médio e superior na Polônia. Comparando diferentes formas de apresentar padrões sintáticos verbais, a pesquisadora concluiu que os códigos gramaticais eram usados apenas pelos estudantes universitários (alunos de Linguística e Gramática Inglesa), mas que ainda assim eles preferiam descrições gramaticais explícitas ao invés de códigos abreviados. Tal como no experimento anterior, realizado na Holanda, Dziemianko constatou que também na Polônia as definições do tipo COBUILD não eram usadas para consultar sintaxe verbal e que os exemplos foram novamente considerados a melhor opção.

Tabela 1: Diferentes formas de apresentar informações gramaticais em dicionários

Exemplos do verbo approve no sentido de “ <i>considerar alguém/algo bom ou adequado</i> ”		Fonte
Códigos sintáticos	V + <i>of</i>	COBUILD
Informação sintática explícita	approve of	LDOCE
	approve of somebody/something approve of somebody doing something approve of somebody's doing something	OALD
	approve of approve of someone doing something	Macmillan
Padrões sintáticos	<i>If you approve of an action, event, or</i>	COBUILD

incorporados nas definições	<i>suggestion, you like it or are pleased about it; If you approve of someone or something, you like and admire them.</i>	
Alertas sobre erros comuns (e exemplos)	<i>When approve means 'to have a positive feeling towards something or someone', it is followed by of, not by a direct object. X I do not approve violence. ✓ I do not <u>approve of</u> violence</i>	Macmillan
Exemplos	He seemed to approve of my choice. I don't really approve of children wearing make-up.	Macmillan
	She doesn't approve of my friends. He doesn't approve of smoking. I thoroughly approve of what the government is doing.	Cambridge
	Do you approve of my idea? She doesn't approve of me leaving school this year. She doesn't approve of my leaving school this year.	OALD
	Not everyone approves of the festival. You've never approved of Henry, have you?	COBUILD
	Catherine's parents now approve of her marriage. I don't approve of cosmetic surgery.	LDOCE

Em um estudo realizado em Hong Kong, Chan (2011) aplicou a um grupo de 31 aprendizes cantoneses de inglês uma tarefa de avaliação gramatical com consulta a dicionário. Alguns dos verbetes fornecidos aos alunos continham informações gramaticais explícitas, enquanto em outros essas informações haviam sido intencionalmente removidas. Ao relatarem suas experiências, os alunos mais uma vez indicaram que preferem obter informações gramaticais por meio de exemplos. Dziemianko (2012) realizou um novo estudo detalhado e em grande escala com 893 estudantes universitários e do ensino médio na Polônia, comparando a reação dos alunos a exemplos e outras formas de se apresentarem informações gramaticais nos dicionários. Mais uma vez, a pesquisadora constatou que os exemplos superaram todas as outras fontes de informações sintáticas.

1.2 Estudos sobre exemplos em dicionários

Apesar de os alunos de línguas se beneficiarem das frases exemplos que ilustram o uso contextualizado das palavras, os primeiros estudos sobre o efeito dos exemplos nos dicionários não foram conclusivos (Summers 1988, Laufer 1993, Nesi 1996, Al-Ajmi 2008). Segundo Frankenberg-Garcia (2012, 2014), uma possível explicação seria que os dicionários não distinguem entre os exemplos que facilitam a compreensão e aqueles que auxiliam na produção oral ou escrita. De acordo com Humblé (2001), os exemplos que facilitam a compreensão (exemplos de decodificação) devem conter pistas contextualizadas sobre o significado das palavras, enquanto que os exemplos que facilitam a produção (exemplos de codificação) devem apresentar padrões gramaticais e colocacionais. O exemplo (1), a partir da palavra-entrada **feign** na versão *online* do *Oxford Advanced Learner's Dictionary* (OALD), contém pistas contextuais como *survived the massacre* e *death*, que podem ajudar o aluno a compreender o significado de **feign**. Como o uso desse verbo transitivo em contexto não oferece grandes dificuldades, pode-se afirmar que (1) é principalmente um exemplo de decodificação. Em contrapartida, o exemplo de dicionário (2), a partir da entrada **cater** na versão *online* do *Macmillan English Dictionary* (MED), ilustra que o verbo é intransitivo e seguido da preposição *for*, mas não parece muito útil no sentido de ajudar os usuários a entender o significado de **cater**. Pode-se afirmar, portanto, que (2) é principalmente um exemplo de codificação. Existem, obviamente, exemplos que ajudam tanto na compreensão quanto na produção linguística. No exemplo (3), retirado do verbete **believe** no MED, *previous supporters* combinado com *stopped* dá uma ideia do significado de **believe**, e o exemplo também mostra que, com esse sentido, o verbo é intransitivo e seguido da preposição *in*.

- (1) He survived the massacre by feigning death.
- (2) I'm catering for a wedding reception next week.
- (3) Some of his previous supporters have stopped believing in him.

No entanto, conforme observado em Frankenberg-Garcia (2012), não parece haver muitos exemplos em dicionários que forneçam pistas contextuais claras sobre o significado da palavra, e que ao mesmo tempo ajudem com a gramática. Entretanto, isso não deve ser considerado uma deficiência dos dicionários, mas sim uma consequência do simples fato de que as pessoas não fornecem pistas contextuais sobre o significado das palavras todas as vezes em que as palavras são utilizadas, o que pode ser comprovado por meio de uma breve análise de um conjunto aleatório de linhas de concordância em um *corpus*. Como mostrado no exemplo (4), as cinco primeiras linhas de concordância do advérbio **deliberately** no *corpus* de língua inglesa SKELL² (Baisa e Suchomel 2014) escolhidas aleatoriamente sugerem que, em geral, esse advérbio é usado no contexto de uma construção passiva. No entanto, as concordâncias não fornecem pistas contextuais muito claras sobre o que **deliberately** significa.

(4) Therefore health issues are **deliberately** brushed aside.
The advertising rules surrounding children are **deliberately** strict.
The word "consultation" is **deliberately** used here.
This dire situation has been **deliberately** created.
Its distribution was **deliberately** delayed three times.

Voltando aos estudos mencionados acima sobre o (pouco) efeito dos exemplos (Summers 1988, Laufer 1993, Nesi 1996, Al-Ajmi 2008), o fato de esses pesquisadores se basearem em exemplos reais de dicionários - como era, de fato, a intenção - significa que não foi possível comprovar se a influência dos exemplos na compreensão linguística foi testada com exemplos de decodificação, e se a influência dos exemplos na produção linguística foi testada com exemplos de codificação. Como discutido em Frankenberg-Garcia (2012, 2014), se o objetivo for testar a eficiência dos exemplos como forma de auxiliar na compreensão, os exemplos fornecidos aos alunos devem ao menos ter esse potencial. Se os exemplos se restringirem a ilustrar a sintaxe ou a colocação na língua estrangeira, sem fornecer pistas contextuais que ajudem a transmitir o significado, é pouco provável que sejam muito úteis para a compreensão. Da mesma forma, exemplos que almejam a produção linguística devem ter o potencial de auxiliar na produção. Se não conseguirem exibir os padrões léxico-gramaticais específicos necessários para satisfazer as necessidades de codificação do aluno, esses exemplos não serão úteis para a produção.

Outra deficiência dos experimentos anteriores com exemplos em dicionários é que os procedimentos de elicitación dos estudos que examinam exemplos e produção linguística (Summers, 1988, Laufer, 1993, Nesi, 1996) geralmente solicitavam que os participantes escrevessem frases com palavras que desconheciam. Assim, os sujeitos provavelmente tiveram que procurar o significado dessas palavras antes que pudessem empregá-las na escrita e, como Summers (1988) e Laufer (1993) reconheceram, não é natural utilizar na produção textual palavras desconhecidas, das quais é necessário procurar o significado. Frankenberg-Garcia (2014) argumentou que nesses experimentos os problemas de compreensão podem ter constituído uma importante variável interveniente, distorcendo os resultados obtidos em relação ao efeito (ou a falta dele) dos exemplos sobre a produção linguística.

Mais recentemente, Frankenberg-Garcia (2012) e Frankenberg-Garcia (2014) revisitou a questão do valor dos exemplos na compreensão e produção linguísticas por meio da realização de alguns experimentos em que, no lugar de recorrerem a exemplos autênticos de dicionários, os participantes (48 alunos de graduação e 50 alunos do ensino médio que estudavam inglês em Portugal) utilizaram exemplos selecionados de *corpora* para garantir que os participantes recebessem exemplos de decodificação nos testes de compreensão e exemplos de codificação nos testes de produção. Mais uma vez contrapondo-se aos estudos anteriores, os testes de produção não apresentavam palavras desconhecidas para os participantes. Ao contrário, as palavras testadas eram razoavelmente fáceis de serem compreendidas (a fim de controlar a variável de

compreensão), mas em geral usadas de forma incorreta por alunos portugueses de inglês (a fim de encorajar os participantes a pesquisar como essas palavras eram usadas). Além disso, como os alunos poderiam necessitar de mais do que apenas um exemplo para conseguir deduzir o significado das palavras ou para aprender a empregá-las de forma gramatical, Frankenberg-Garcia (2012, 2014) também comparou as diferenças entre fornecer aos alunos um único ou múltiplos exemplos.

Os dois estudos apresentaram resultados muito semelhantes em relação aos efeitos dos exemplos sobre a compreensão: um exemplo de decodificação ajudou um pouco os alunos a entenderem as palavras desconhecidas, e múltiplos exemplos de decodificação funcionaram tão bem quanto as definições de dicionário. Embora ainda seja necessário replicar esses estudos, o presente trabalho não trata de exemplos para a compreensão linguística, pois essas informações já existem nas definições fornecidas em dicionários monolíngues e nas traduções dos dicionários bilíngues.

1.3 Exemplos de codificação

Por outro lado, para a produção linguística os exemplos permanecem cruciais. Apesar de os códigos sintáticos, ou as classes gramaticais apresentadas de forma explícita, assim como as informações sintáticas incorporadas em definições do tipo COBUILD serem úteis, parece que os exemplos não só são mais utilizados, como são mais eficazes. Em Frankenberg-Garcia (2012), tanto um único exemplo quanto múltiplos exemplos de codificação ajudaram os participantes a cumprirem a tarefa de forma bem mais adequada do que o grupo de controle, e os participantes com acesso a múltiplos exemplos de codificação superaram significativamente aqueles que tiveram acesso a apenas um exemplo. No entanto, em Frankenberg-Garcia (2014), apenas os participantes que utilizaram múltiplos exemplos de codificação obtiveram escores significativamente mais altos do que os do grupo de controle. Como discutido em Frankenberg-Garcia (2014), essa discrepância pode ter decorrido das diferentes formas de elicitar os dados nos dois estudos. Enquanto no estudo de 2012 os alunos receberam um teste de correção de erros e, portanto, estavam cientes de que as palavras em destaque haviam sido mal empregadas, no estudo de 2014 os dados foram levantados por meio de uma tarefa de tradução. Para essa tarefa, os participantes não foram informados que as palavras testadas eram, em geral, problemáticas. Em outras palavras, os participantes do estudo de 2014 podem ter presumido que as palavras que precisaram empregar na tarefa de tradução eram bastante óbvias e podem não ter consultado os exemplos que lhes foram fornecidos no teste de forma tão cuidadosa quanto os participantes do estudo de 2012, que receberam orientações explícitas para tentarem corrigir erros. A diferença entre os dois procedimentos de elicitação não deve ser menosprezada. Não há razão para que os alunos consultem quaisquer fontes de referências se acharem que já sabem como usar o idioma, ainda que estejam equivocados a esse respeito. Como discutido em Frankenberg-Garcia (2005, 2011), uma das principais razões pelas quais os alunos não usam os dicionários tanto quanto deveriam é que eles, muitas vezes, não estão cientes das suas próprias

limitações linguísticas e necessidades de consulta. Assim, não importa se os dicionários apresentam as informações sobre classes gramaticais de forma adequada. Os alunos não irão consultá-las se não considerarem necessário.

Resumindo, o que sabemos até agora é que, embora os exemplos não sejam a única fonte de informação gramatical para ajudar os alunos com a produção linguística, em geral os alunos preferem recorrer a exemplos e parecem aprender mais com exemplos do que com códigos sintáticos, ou a partir de informações gramaticais explícitas, ou de definições como as encontradas no COBUILD. Além disso, exemplos que exibem a sintaxe da língua estrangeira parecem ser bastante eficazes, especialmente se os alunos estiverem cientes de suas limitações linguísticas e se tiverem à disposição mais de um exemplo.

Embora o espaço seja uma das principais limitações dos dicionários impressos, agora que os dicionários eletrônicos estão se tornando cada vez mais populares, o número de exemplos fornecidos não terá um grande impacto em termos de custos. Na verdade, em algumas das versões eletrônicas dos principais dicionários de inglês para estrangeiros, alguns verbetes possuem hiperlinks para exemplos adicionais. Mas será que os alunos consultam esses exemplos adicionais? E se o fazem, serão esses exemplos adicionais úteis? Muitos exemplos podem confundir ou mesmo sobrecarregar os aprendizes, e essa sobrecarga de informação pode ser prejudicial para a aprendizagem. Além disso, não é apenas a quantidade de exemplos que importa, mas também os tipos. Quando clicamos no "banco de exemplos" do CD-ROM do LDOCE 4 para procurar **approve**, por exemplo, as frases apresentadas não estão organizadas de acordo com as diferentes acepções - *concordar com algo* e *achar que alguém/algo é aceitável ou adequado* -, o que pode ser importante para fins de decodificação. Também não são classificadas quanto à transitividade, o que pode ser importante para fins de codificação. Conforme mostrado em (5), um aluno que necessite decodificar pode se confundir com a mistura de acepções, e um aluno que precise codificar pode se confundir com a mistura de empregos do verbo como transitivo e intransitivo dos primeiros cinco exemplos apresentados no "banco de exemplos adicionais" do LDOCE 4. Isso significa que, sem uma orientação correta, o "banco de exemplos" pode ser tão difícil de interpretar quanto os dados brutos de um *corpus*.

(5) Bob doesn't **approve** of alcohol.

Congress voted not to **approve** the President's plan for cutting the arms budget.

Do you **approve** of doctors using human embryos for research?

Don will only buy the motorcycle if his parents **approve**.

Forty percent of registered voters **approved** of Senator Campbell.

Na seção dedicada à discussão em Frankenberg-Garcia (2014), argumento que, ao apresentar múltiplos exemplos aos alunos, é importante garantir que esses exemplos sejam do mesmo tipo. Embora os exemplos de decodificação devam ser classificados por significado, isto é, de acordo com as diferentes acepções das palavras polissêmicas, os

exemplos de codificação precisam ser organizados de acordo com os diferentes padrões léxico-gramaticais de uso. Além disso, também recomendo que haja exemplos suficientes de cada tipo, de modo que os alunos sejam repetidamente expostos a esses padrões.

1.4 Objetivos do estudo

Tendo em vista o contexto e a discussão apresentados acima, este trabalho tem como objetivo investigar mais a fundo se a exposição aos exemplos de codificação ajudaria os alunos na produção linguística. O procedimento de elicitación foi refinado para garantir que os alunos não fizessem um simples teste de correção de erros, e que, ao mesmo tempo, fossem motivados a consultar os exemplos fornecidos, mesmo que o teste de produção linguística utilizado para elicitar os dados parecesse inicialmente muito fácil. Além disso, busquei investigar se haveria um número ideal de exemplos a serem fornecidos aos alunos. Em Frankenberg-Garcia (2012, 2014), os alunos receberam apenas um ou exatamente três exemplos. Mas os alunos poderiam se beneficiar de dois exemplos? E quatro exemplos ajudariam mais? Neste estudo, portanto, também comparei os efeitos de um, dois, três e quatro exemplos.

2. Método

2.1 Sujeitos

Participaram do estudo 161 alunos de graduação de uma universidade portuguesa. Diferentemente de vários outros estudos sobre o uso de dicionários, os participantes não eram estudantes de Letras ou Linguística, mas graduandos em Turismo (49 estudantes), Hotelaria (87 estudantes), Administração de Empresas (4 estudantes), Sistemas de Informação para Negócios (7 alunos), Secretariado (4 alunos), Marketing (6 alunos) e Recursos Humanos (4 alunos). Todos os alunos tinham duas horas semanais de aula de Inglês, que valiam créditos. Egressos do sistema educacional de ensino secundário português, em que a língua inglesa é obrigatória até o 11º ano e facultativa no 12º, os alunos que participaram do estudo foram aqueles matriculados no nível 4 da cadeira de Inglês da universidade, o que equivale aos níveis B1 e B2 (CEFR).⁴ A média de idade dos sujeitos era 22,7 anos. Desses, 64 eram do sexo masculino e 96 do sexo feminino. Um aluno não preencheu o campo "gênero".

2.2 Procedimento de elicitación

O teste aplicado aos participantes para extrair os dados para o presente estudo foi realizado durante as aulas de inglês. Para a tarefa, os alunos foram agrupados de acordo com o nível de proficiência em língua inglesa e período de estudo (manhã, tarde ou noite), e não com os cursos universitários. O teste foi conduzido por quatro professores diferentes, responsáveis por oito grupos de inglês nível 4 no mesmo turno da tarde. A maioria dos alunos desse turno estuda Hotelaria ou Turismo, o que explica por que havia menos alunos de outras disciplinas realizando o teste.

Os professores responsáveis por administrar o teste receberam instruções escritas sobre como proceder e as instruções também foram discutidas oralmente para garantir

que não restassem dúvidas sobre as rigorosas disposições necessárias para a condução do experimento. Como os professores normalmente falavam apenas em inglês na sala de aula, o teste foi ministrado em inglês, mas dúvidas poderiam ser esclarecidas em português.

O experimento foi realizado como em ambiente de prova, mas sem limite de tempo. Levou entre cinco e dez minutos para que os alunos preenchessem o pré-teste, e outros cinco a dez minutos para completarem o pós-teste realizado na mesma sessão. Ambos serão descritos a seguir.

2.2.1 Tarefa de tradução pré-teste.

Na primeira parte do experimento, o pré-teste, os participantes deveriam traduzir algumas sentenças para o inglês. Essa tarefa poderia induzi-los a cometer erros típicos de falantes nativos de português. Nessa etapa do teste, não foi permitido o uso de qualquer fonte de consulta, e os alunos deveriam escrever as respostas com caneta verde, por razões que seriam reveladas na fase pós-teste. Os dez itens do pré-teste e os erros típicos associados a eles encontram-se na Figura 1. Como pode ser observado, todas as traduções foram bastante literais quanto ao vocabulário em inglês, mas, ao mesmo tempo, exigiam que os alunos empregassem palavras que geralmente causam dificuldade nos contextos apresentados⁵. Vejamos, por exemplo, o item 1 do teste. Os estudantes portugueses de nível B1/B2 (CEFR) normalmente não têm problemas com o léxico necessário para traduzir o trecho em questão. Contudo, devido à influência do português, eles geralmente cometem o erro **it's the first time someone does something* (em vez de *it's the first time someone has done something*). Embora os alunos também possam cometer outros tipos de erros ao traduzir essas frases – como erros de ortografia, ou a tradução de *à noite* no item 7 do teste para *at night*, em vez de *tonight*, como será visto mais adiante na seção 2.2.5 –, este estudo se limitou a avaliar os erros específicos destacados na Figura 1.

1. É a primeira vez que venho a este restaurante.
It's the first time I have come/been to this restaurant. ✓
It's the first time I come to this restaurant. ✗
2. A piscina tem vinte metros de comprimento.
The swimming pool is twenty metres long. ✓
The swimming pool has twenty metres. ✗
3. Aprova o trabalho do presidente?
Do you approve of the president's work? ✓
Do you approve the president's work? ✗
4. Ele trabalha connosco desde 2010.
He has worked/has been working with us since 2010. ✓
He works with us since 2010. ✗
5. Eles nunca confiaram em nós.
They (have) never trusted us. ✓
They never trusted in us. ✗
6. Eu sempre segui os seus conselhos.
I (have) always followed your advice. ✓
I always followed your advices. ✗
7. Ligue-me hoje à noite para o 212 745 975.
Call me tonight at/on 212 745 975. ✓
Call me tonight to 212 745 975. ✗
8. Parabéns pelo seu novo emprego.
Congratulations on your new job. ✓
Congratulations for your new job. ✗
9. Telefone para o banco amanhã de manhã.
Telephone the bank tomorrow morning. ✓
Telephone to the bank tomorrow morning. ✗
10. Votei no candidato errado.
I voted for the wrong candidate. ✓
I voted in the wrong candidate. ✗

Fig. 1: Tarefa parcial de tradução, traduções corretas e erros típicos previstos

2.2.2 Grupos experimentais.

Depois de concluírem a tarefa de tradução do pré-teste descrita acima, os participantes devolveram as canetas verdes que receberam para realizar as traduções e foram aleatoriamente divididos em cinco grupos diferentes em cada sala de aula. O grupo 1 (totalizando 32 alunos) recebeu como material de referência uma folha contendo um exemplo de codificação por item do teste; o grupo 2 (33 alunos) recebeu uma folha com dois exemplos de codificação por item do teste; o grupo 3 (32 alunos) obteve três exemplos de codificação; e o grupo 4 (33 alunos), quatro exemplos de codificação por item do teste. O grupo 5 (31 alunos) era o grupo controle e, portanto, não recebeu material de consulta.

2.2.3 Exemplos de codificação.

Os exemplos de codificação fornecidos aos participantes foram selecionados a partir de *corpora* de língua inglesa geral (o *British National Corpus*, o *Corpus of Contemporary American English* e o *en Ten Ten*).⁶ Algumas concordâncias foram reduzidas para que não ultrapassassem uma linha. Os exemplos de codificação deveriam se

restringir à sintaxe da língua estrangeira presente no teste. Por exemplo, as concordâncias em (6) foram selecionadas para mostrar que o presente perfeito normalmente é usado após a expressão *it's the first time* no item 1 do teste.

(6) **It is the first time** she's ever been here.

It is the first time we have attended the exhibition.

This is a brand new idea, **it is the first time** I have presented it in public.

If **it is the first time** you have had the symptoms, consult your doctor.

O procedimento de seleção buscou garantir que os exemplos escolhidos fossem facilmente compreendidos pelos alunos, com 95% do vocabulário contido na lista *Oxford 3000*. As palavras fora da lista *Oxford 3000* ou eram nomes próprios ou possuíam cognatos em português. Vale ressaltar, no entanto, que em alguns casos o conjunto de exemplos fornecidos não era totalmente simétrico. Por exemplo, conforme destacado em (7), para o item 6 do teste diferentes estruturas foram utilizadas para ilustrar que a palavra *advice* é incontável: *a lot of advice*, *words of advice*, sem quantificador, e *pieces of advice*. Já que nos quatro casos o equivalente em português estaria no plural, todos eles possibilitaram que os alunos percebessem que não se pode dizer **advices* em inglês. Embora os participantes pudessem deparar com esses usos da palavra também em situações reais de comunicação, experimentos futuros talvez deveriam garantir que os exemplos fornecidos aos diferentes grupos sejam o mais comparável possível.

(7) She gave him a lot of **advice** and help.

Thank you so much for your words of **advice**.

The first step is do not be afraid to seek **advice**.

I was given three pieces of **advice**: don't write such complicated sentences, slow down and smile.

A folha de referência com o conjunto completo de concordâncias fornecidas ao grupo experimental com quatro exemplos encontra-se no Apêndice. O grupo experimental que recebeu um único exemplo teve acesso apenas à primeira concordância de cada conjunto de quatro dessa lista; o grupo experimental que usou dois exemplos recebeu as duas primeiras concordâncias; e o grupo experimental que usou três exemplos recebeu as três primeiras concordâncias.

2.2.4 Tarefa de revisão pós-teste.

Depois que as respectivas folhas de referência foram distribuídas aos diferentes grupos experimentais, os alunos foram informados de que as frases que tinham acabado de traduzir podiam parecer fáceis, mas na verdade eram bastante problemáticas, já que os falantes nativos de português geralmente cometem erros ao transportá-las para o inglês. Os participantes foram então solicitados a iniciar o pós-teste. Nessa etapa, os alunos

deveriam revisar as traduções que haviam acabado de escrever com a caneta verde, e, se achassem que haviam cometido algum erro, deveriam refazer a tradução embaixo, usando as próprias canetas, azuis ou pretas. A mudança de cor visava garantir que os alunos ainda pudessem ver o que escreveram na fase de pré-teste, mas não se sentissem tentados a escrever por cima dessas traduções. Além disso, dessa forma quaisquer modificações feitas do pré-teste para o pós-teste ficariam claramente visíveis. Os alunos do grupo de controle revisaram as traduções sem consulta, mas, assim como os outros grupos, eles foram informados de que as traduções na fase de pré-teste podiam ser menos simples do que pareciam. Os alunos restantes foram informados de que as folhas de consulta que receberam continham frases que poderiam ajudá-los a melhorar suas traduções e foram solicitados a ler essas frases cuidadosamente enquanto revisavam as traduções⁷. Foi ainda explicado aos alunos que as frases sob o número 1 nas folhas de referência poderiam ajudar com o item 1 do teste, que aquelas sob o número 2 poderiam ajudar no item 2 do teste, e assim por diante. Isso foi feito porque, da mesma forma que no estudo de Bogaards e van der Kloot (2002), o foco do presente estudo era avaliar a eficácia dos exemplos, e não se eram fáceis de encontrar.

O formato pré-teste e pós-teste apresentado foi uma tentativa de refinar os procedimentos de eliciação adotados anteriormente em Frankenberg-Garcia (2012) e Frankenberg-Garcia (2014). Como descrito na Introdução, no estudo de 2012 os indivíduos receberam um teste de correção de erros, o que significa que eles deveriam corrigir erros que não foram cometidos por eles próprios, comprometendo, até certo ponto, a validade ecológica do estudo. No estudo de 2014, embora os participantes tenham recebido uma tarefa de tradução semelhante à usada no presente estudo, eles não foram informados de maneira explícita que as frases aparentemente fáceis que iriam traduzir eram, na verdade, propensas a erros. Como discutido anteriormente, isso significa que alguns dos participantes mais confiantes podem não ter consultado os exemplos fornecidos durante o teste, ou não com a devida atenção, o que pode ter afetado os resultados. No presente experimento, houve o cuidado de assegurar que os participantes fossem levados a reavaliar as próprias traduções iniciais (em vez de avaliar erros que poderiam não ter sido cometidos por eles), e encorajados a usar os exemplos fornecidos (em vez de ingenuamente dispensá-los porque o teste parecia fácil).

2.2.5 Corrigindo as traduções.

Todos os testes foram corrigidos por um único avaliador, sem uma ordem definida. Para garantir consistência, os testes foram corrigidos item por item em vez de participante por participante. Os detalhes pessoais de cada participante e, mais importante, as informações sobre o grupo experimental ao qual pertenciam, constavam na folha de rosto do teste, que não estava visível para o avaliador durante o procedimento.

Cada estrutura usada corretamente na língua estrangeira valia um ponto. Como mostrado na Figura 1, foram aceitas variações nas traduções, inclusive aquelas que não estavam previstas na Figura 1, como *the president's work* vs. *the work of the president* no

item 3 do teste, ou *tomorrow morning* vs. *tomorrow during the morning* no item 9. Erros ortográficos que não estivessem relacionados às estruturas sendo analisadas, ou escolhas lexicais inadequadas, como *your new work* em vez de *your new job* no item 8, foram desconsiderados. Os erros identificados, no entanto, foram na maior parte exatamente aqueles previstos na Figura 1. Correções parciais receberam meio ponto. Esse caso ocorreu apenas com o item 2 do teste, em que alguns dos estudantes acrescentaram a palavra *long* ao final **This swimming pool has twenty metres*, mas não substituíram o verbo *has* por *is*. Embora a sentença **This swimming pool has twenty metres long* também esteja incorreta, nesse caso os alunos ao menos conseguiram perceber a presença do advérbio no final da construção. Por isso, o meio acerto.

Como os testes não foram devolvidos aos alunos, não houve tentativa de fornecer *feedback* corretivo, embora tenha sido permitido que os professores responsáveis pela aplicação dos testes fornecessem aos alunos as respostas corretas na aula seguinte, caso desejassem.

Cada teste recebeu duas notas. A nota do pré-teste considerou as traduções iniciais, em caneta verde. A nota do pós-teste considerou as traduções da segunda etapa, depois que os participantes foram encorajados a revisar as primeiras traduções com a ajuda das folhas de consulta fornecidas aos grupos experimentais ou, no caso do grupo de controle, sem consulta. Esse procedimento permitiu a comparação das notas do pré-teste e do pós-teste de cada aluno individualmente e, em seguida, a comparação das diferenças gerais de pontuação entre grupos. Os resultados obtidos serão apresentados na próxima seção.

3. Resultados

Os resultados preliminares do presente estudo estão resumidos nas Tabelas 2a e 2b. As pontuações do pré-teste na Tabela 2a mostram que a tarefa de tradução parcial usada foi bem difícil para a maior parte dos alunos, resultando em uma pontuação média geral abaixo de 2, e a mediana foi de apenas 1 ponto de 10. Esse resultado foi um bom indicador de que as questões do teste eram suficientemente difíceis para que houvesse potencial de aperfeiçoamento na fase pós-teste. Houve, no entanto, um pequeno número de alunos que teve um desempenho muito melhor do que os outros no pré-teste, como pode ser observado nos números apresentados na coluna “intervalo”. As pontuações do pré-teste de cada grupo indica que não houve grande diferença entre os grupos, o que reflete a maneira aleatória como os alunos foram designados para cada grupo. Note, no entanto, que um aluno do grupo de controle com pontuação 9, dois alunos do grupo de 2 exemplos com pontuação 8 e 9 e um aluno do grupo de 3 exemplos com pontuação 10 foram considerados casos isolados. Seria muito difícil melhorar o desempenho desses alunos durante a etapa de pós-teste, uma vez que os escores obtidos nos pré-testes foram excepcionalmente altos.⁸ Assim, decidiu-se excluir esses casos isolados da análise mais aprofundada. Por terem tido desempenho alto demais no pré-teste, esses alunos não se qualificariam como candidatos adequados para o presente experimento. Além disso, a exclusão de quatro dos 161 participantes selecionados inicialmente não representaria uma

grande perda para o experimento. Após a exclusão desses quatro participantes, os valores das pontuações do pré-teste foram recalculados e são apresentados na Tabela 2b.

As pontuações do pós-teste são, por sua vez, apresentadas na Tabela 3, que mostra que as pontuações média e mediana dos quatro grupos experimentais aumentaram, ao passo que houve apenas um ligeiro aumento na média e nenhuma mudança na pontuação mediana do grupo de controle. Curiosamente, houve também um aumento nos desvios-padrão e intervalos para todos os grupos, exceto o grupo de controle, indicando que os exemplos de codificação fornecidos aos grupos experimentais podem ter acentuado as diferenças individuais entre os alunos.

A Tabela 4 resume as diferenças entre as pontuações pré e pós-teste, e evidencia que, em média, os participantes melhoraram quase três pontos nos pós-testes. No entanto, o grupo de controle melhorou muito pouco ou nada (com média de 0,23 e mediana de 0), enquanto que o grupo de 3 exemplos foi o que teve o maior aproveitamento, com uma diferença média de 4,13 e uma mediana de 4. Note, contudo, que houve bastante alteração dentro dos grupos, com mudanças gerais nas pontuações variando de -2 a 8.5.

Tabela 2a: Pontuações preliminares no pré-teste

Grupo	N	Média da nota no pré-teste (de 0 a 10)	Desvio-padrão	Intervalo	Casos isolados	Mediana
Controle	31	2,00	2,37	0-9	1	1
1 exemplo de codificação	32	1,44	1,56	0-5	0	1
2 exemplos de codificação	33	2,18	2,26	0-9	2	2
3 exemplos de codificação	32	2,16	2,30	0-10	1	1
4 exemplos de codificação	33	1,70	2,05	0-7	0	1
Todos os grupos	161	1,90	2,21	0-10	4	1

Tabela 2b: Pontuações no pré-teste (após exclusão dos casos isolados)

Grupo	N	Média da nota no pré-teste (de 0 a 10)	Desvio-padrão	Intervalo	Mediana
Controle	30	1,77	2,01	0-7	1

1 exemplo de codificação	32	1,44	1,56	0-5	1
2 exemplos de codificação	31	1,77	1,61	0-6	1
3 exemplos de codificação	31	1,90	1,83	0-7	1
4 exemplos de codificação	33	1,73	2,05	0-7	1
Todos os grupos	157	1,72	1,81	0-7	1

Para se ter uma ideia mais clara das diferenças entre as pontuações pré e pós-teste para cada grupo separado e avaliar se alguma melhora foi significativa, os dois conjuntos de dados foram comparados usando o teste Wilcoxon. A escolha desse teste, em vez do teste-t, mais comum para dados pareados, deve-se ao fato de que os dados não apresentavam uma distribuição normal. As hipóteses nulas eram de que, para cada grupo, não haveria melhora do pré-teste para o pós-teste. Os resultados dessa análise estão resumidos na Tabela 5, que mostra que as hipóteses nulas podem ser rejeitadas para todos os grupos experimentais com um grau de confiança de 95%, mas não para o grupo de controle. A Tabela 5 também mostra que a maioria dos 30 participantes designados para o grupo de controle não pontuou melhor no pós-teste, e cinco deles obtiveram notas mais baixas após a revisão das traduções iniciais. Em contrapartida, em todos os grupos experimentais a maioria dos participantes pontuou melhor no pós-teste. Poucos não conseguiram melhorar as traduções, e nenhuma tradução teve a nota reduzida. Note, no entanto, que houve alguns casos em que os alunos conseguiram corrigir a tradução de um item do teste, mas depois introduziram um erro em outro item.

Tabela 3: Pontuações no pós-teste

Grupo	N	Média da nota no pós-teste (de 0 a 10)	Desvio-padrão	Intervalo	Mediana
Controle	30	2,00	1,98	0-7	1
1 exemplo de codificação	32	4,53	2,59	0-9	4
2 exemplos de codificação	31	5,32	2,58	0-10	5,5
3 exemplos de codificação	31	6,03	2,61	0-10	7

4 exemplos de codificação	33	5,18	2,20	1,5-9	5
Todos os grupos	157	4,64	2,74	0-10	4,5

Tabela 4: Diferenças entre as pontuações no pré-teste e no pós-teste

Grupo	N	Média da diferença entre as pontuações no pré-teste e no pós-teste	Desvio-padrão	Alcance	Mediana
Controle	30	0,23	1,16	-2-4	0
1 exemplo de codificação	32	3,19	2,21	0-8	3
2 exemplos de codificação	31	3,55	2,26	0-8	3
3 exemplos de codificação	31	4,13	2,21	0-8,5	4
4 exemplos de codificação	33	3,45	1,69	0,5-7	3,5
Todos os grupos	157	2,49	2,36	-2-8,5	2,5

Tabela 5: Diferenças entre as pontuações no pré-teste e no pós-teste (teste Wilcoxon)

Grupo	N	Pontuação mais alta no pós-teste	Pontuação mais baixa no pós-teste	Pontuação igual no pós-teste	Valor de Z*	Valor de W*
Controle	30	7	5	18	1,0198	26
1 exemplo de codificação	32	28	0	4	4,6226*	0*
2 exemplos de codificação	31	30	0	1	4,7821*	0*
3 exemplos de codificação	31	29	0	2	4,7030*	0*
4 exemplos de	33	33	0	0	5,0119*	0*

codificação						
--------------------	--	--	--	--	--	--

* = significativo em $p \leq 0,05$

Depois de analisar as diferenças nas pontuações pré-teste e pós-teste para cada grupo separadamente, o passo seguinte foi comparar os diferentes grupos entre si. Para realizar essa comparação, foi utilizado o teste não paramétrico de Kruskal-Wallis, em vez da análise de variância (ANOVA) unidirecional, dada a falta de distribuição normal. A hipótese nula foi de que não haveria diferença entre os grupos. Com nível de significância de 0,05 e grau de liberdade de 4 (dados os cinco grupos sendo comparados), observou-se que o valor crítico de X^2 (9,48773) é menor do que o valor calculado H (55,986), o que significa que a hipótese nula pode ser rejeitada.⁹ Pode-se concluir, portanto, que existem diferenças significativas entre os cinco grupos. Analisando as diferenças reais nas pontuações do pré e do pós-teste na Tabela 4, parece natural atribuir a principal causa dessa diferença ao grupo de controle, que não recebeu nenhum tratamento experimental.

O resultado acima não esclarece, entretanto, se o número de exemplos apresentados aos participantes de cada grupo realmente fez diferença. Portanto, para focar as diferenças entre o número de exemplos, o teste de Kruskal-Wallis foi usado novamente, mas dessa vez deixando intencionalmente de fora o grupo de controle. A hipótese nula era de que não haveria diferença entre os tratamentos (ou seja, a exposição dos participantes a um, dois, três ou quatro exemplos). Observou-se que o valor crítico de X^2 em nível 0,05 com grau de liberdade de 3 (7,81473) é maior do que o valor H calculado (3,713), o que significa que a hipótese nula não pode ser rejeitada. Portanto, não é possível concluir que as diferenças entre os diferentes grupos experimentais sejam significativas. Assim, apesar da melhora ligeiramente maior no pós-teste do grupo de 3 exemplos na Tabela 4, essa diferença não foi estatisticamente significativa.

4. Discussão

No geral, os resultados apresentados na seção anterior são condizentes com as conclusões de Frankenberg-Garcia (2012, 2014) e, como em Bogaards e van der Kloot (2002), Chan (2011) e Dziemianko (2006 e 2012), confirmam a hipótese de que os exemplos realmente parecem auxiliar na produção linguística. Mais especificamente, no presente estudo foi possível mostrar que simplesmente advertir os aprendizes que as traduções estão propensas a erros não melhora, por si só, os resultados, ao passo que parece ser proveitoso fornecer exemplos que os ajudem na revisão das primeiras traduções. Na verdade, as pontuações mais baixas no pós-teste de alguns participantes do grupo de controle indicam que apenas chamar a atenção dos alunos para áreas problemáticas pode até ser contraproducente. Naturalmente, expor os alunos a um exemplo qualquer não seria de grande valia. Diferentemente de estudos anteriores, que se concentraram em exemplos de dicionários (cf. Summers 1988, Laufer 1993 e Nesi 1996), e dos exemplos existentes em dicionários para aprendizes, que não distinguem entre exemplos de codificação e decodificação, os exemplos utilizados no presente estudo e em Frankenberg-Garcia (2012,

2014) foram escolhidos criteriosamente para garantir que ilustrassem a sintaxe ou a colocação necessária.

Depois de verificar que os exemplos de codificação que focam a sintaxe ou a colocação na língua estrangeira são positivos, resta o desafio de garantir que os usuários de dicionários consigam acessar exemplos adequados de codificação sempre que necessário. O presente estudo (e Frankenberg-Garcia 2012, 2014) foi realizado em um ambiente experimental controlado, onde os participantes receberam exemplos apropriados. No mundo real, entretanto, os estudantes precisariam encontrar esses exemplos por si mesmos. Uma possível solução seria que os alunos retirassem os próprios exemplos de *corpora*. No entanto, esses exemplos não seriam classificados de acordo com o significado ou o uso, e lidar com dados brutos de *corpus* pode ser bastante assustador no início.¹⁰ Outra possibilidade seria que os dicionários de aprendizes organizassem os exemplos de codificação para fins de produção linguística. Mas ainda que isso fosse feito, as pesquisas apontam que os usuários continuam a consultar os dicionários principalmente para compreensão da linguagem, e muitos dos recursos dos dicionários que auxiliam na produção da linguagem permanecem subexplorados, ou seja, são deixados em segundo plano quando comparados às informações que auxiliam na compreensão. É, portanto, preciso desenvolver formas eficientes de disponibilizar exemplos adequados de codificação aos aprendizes.

O outro objetivo principal do presente estudo foi tentar determinar se o número de exemplos fornecidos tinha impacto na tradução. A partir dos resultados obtidos aqui, não se pode concluir que o número de exemplos de codificação fornecidos tenha feito diferença. Assim, diferentemente de Frankenberg-Garcia (2012), em que três exemplos de codificação ajudaram mais do que apenas um exemplo, e ao contrário de Frankenberg-Garcia (2014), em que um exemplo não levou a melhorias significativas, mas três exemplos sim, no presente estudo foi possível observar que um número maior de exemplos foi mais eficaz do que apenas um exemplo, mas não foi possível determinar um número ideal de exemplos.

Há vários fatores que podem explicar esse resultado. O primeiro ponto a ser discutido é o tamanho da amostra. Com 157 sujeitos (depois que os 4 casos isolados foram excluídos), o presente estudo contou com três vezes mais participantes do que o estudo de 2012 (48 sujeitos), assim como o estudo de 2014 (50 sujeitos). Outra questão relacionada é a própria amostragem. Considerando que os participantes do estudo de 2012 foram recrutados de três turmas do curso de Turismo da mesma universidade e os do estudo de 2014 foram de dois grupos do 11º ano da mesma escola secundária, conforme relatado na seção 2.1.1, neste estudo os participantes eram de grupos mistos de alunos de graduação de cursos de diferentes áreas. O tamanho maior da amostra e as origens mais variadas dos participantes devem contribuir para tornar as descobertas atuais mais generalizáveis do que as dos estudos anteriores.

Além disso, o procedimento refinado de elicitación adotado no presente estudo também permitiu avaliar o progresso do pré-teste ao pós-teste, em vez de avaliar apenas os

resultados pós-tratamento, como foi o caso dos estudos de 2012 e 2014. Embora esses experimentos tenham utilizado grupos de controle para explicar o conhecimento prévio da língua em análise, o procedimento de pré-teste e pós-teste utilizado no presente estudo ajudou a desvendar ainda mais o possível efeito dessa variável – tanto que foi tomada a decisão de excluir da amostra os quatro casos com pontuações superiores a 8 na fase de pré-teste.

Outro fator interessante relacionado ao procedimento de elicitación é que, como discutido na Introdução, no estudo de 2012 os alunos realizaram um teste que focou os erros que eles podem nem cometer ou mesmo notar, enquanto em 2014 os participantes não tinham conhecimento de que as frases que deveriam traduzir eram problemáticas. Em contrapartida, no presente estudo os alunos foram explicitamente solicitados a reavaliar possíveis problemas nas próprias traduções. É, portanto, possível que eles estivessem mais propensos a usar os exemplos fornecidos para encontrar formas de melhorar a própria escrita do que nos estudos anteriores. Isso poderia explicar por que mesmo os participantes com acesso a apenas um exemplo tiveram resultados tão bons quanto os dos participantes que tinham acesso a mais de um exemplo.

As diferenças individuais entre os sujeitos também devem ser levadas em consideração na interpretação dos resultados. Se o número de exemplos apresentados aos alunos não parece fazer diferença no geral, pode ser que alguns indivíduos prefiram e até mesmo reajam melhor quando se deparam com menos exemplos, enquanto outros se beneficiem e necessitem de mais exemplos. Pode até mesmo haver aprendizes para os quais exemplos de codificação não ajudem em nada. De fato, os desvios-padrão e os intervalos observados para todos os grupos, exceto o grupo de controle, na fase pós-teste, indicam que os exemplos fornecidos aos grupos experimentais podem ter acentuado as diferenças individuais. Assim, também é necessário realizar mais pesquisas sobre as preferências e necessidades individuais antes de determinar um número ideal de exemplos.

5. Conclusão

Este estudo surgiu da necessidade de explorar mais a fundo como ajudar os alunos na produção linguística. Apesar de a presença de exemplos nos dicionários ser bem-vinda, não há provas contundentes de que seja suficiente. Parti da hipótese de que para ajudar os aprendizes na produção linguística é importante usar exemplos de codificação (em oposição à decodificação), e organizar esses exemplos de acordo com diferentes padrões léxico-gramaticais de uso. Com base em descobertas anteriores relatadas em Frankenberg-Garcia (2012, 2014), também se faz necessário fornecer aos alunos mais do que um único exemplo. O presente estudo, portanto, investigou se os exemplos de codificação ajudariam os alunos na produção linguística, por meio de um procedimento refinado de elicitación aplicado a grupos experimentais e de controle maiores, e avaliou se os alunos reagiriam melhor a um número ideal de exemplos. Os resultados obtidos reforçam a ideia de que os exemplos de codificação são úteis como auxílio à escrita, mas não foi possível determinar qual seria o número ideal de exemplos, sugerindo que, nesse aspecto, pode haver uma

significativa variabilidade entre os aprendizes. Resta ainda entender melhor as preferências individuais dos alunos em relação à quantidade de exemplos de que necessitam, assim como desenvolver formas mais eficientes de disponibilizá-los.

Notas da autora

1. As informações da Tabela 1 foram extraídas de cinco dicionários de aprendizes, em 03/11/2014: *Collins English for Learners* (COBUILD), *Oxford Advanced Learners' Dictionary* (OALD), *Cambridge Learners Dictionary* (Cambridge), *Macmillan English Dictionary* (Macmillan), *Longman Dictionary of Contemporary English* (LDOCE).
2. Fonte: <http://skell.sketchengine.co.uk/> em 17/11/2014.
3. Além de usar uma amostra populacional diferente, também foram feitas mudanças no design experimental.
4. Quadro Comum Europeu de Referência para Línguas. Ver http://www.coe.int/t/dg4/linguistic/cadrel_en.asp.
5. Todo o repertório de vocabulário necessário para essa tarefa encontra-se na lista Oxford 3000, que contém as 3000 palavras fundamentais da língua inglesa em termos de frequência e uso. Ver <http://oald8.oxfordlearnersdictionaries.com/oxford3000/>.
6. Disponíveis, respectivamente, em <http://corpus.byu.edu/bnc/> (BNC), <http://corpus.byu.edu/coca/> (COCA) e <http://www.sketchengine.co.uk> (en TenTen), acessados em 15/04/2015. Ver também Davies (2008) e Kilgariff (2004) para maiores detalhes sobre esses *corpora*.
7. O termo 'sentenças' foi utilizado deliberadamente em vez de 'concordâncias' para não confundir os participantes com a terminologia da Linguística de Corpus.
8. De acordo com o teste de John Tukey (Tukey, 1977), os casos isolados estão abaixo do primeiro quartil ou acima do terceiro quartil.
9. H = Estatística do teste Kruskal-Wallis.
10. Note, entretanto, que grande progresso já foi feito em relação ao desenvolvimento de interfaces que aproximem leigos de *corpora*. Um exemplo recente e marcante é a interface *corpus-cum-online* do SKELL, do Sketch Engine (<http://skell.sketchengine.co.uk/run.cgi/skell>). Ver também Baisa e Suchomel (2014).

Referências

Dicionários

Cambridge Learners Dictionary Online. Accessed on 3 November 2014. <http://dictionary.cambridge.org/dictionary/learner-english>.

Collins English for Learners (COBUILD) Online. Accessed on 3 November 2014. <http://www.collinsdictionary.com/dictionary/english-cobuild-learners>.

Longman Dictionary of Contemporary English (LDOCE) Online. Accessed on 3 November 2014. <http://www.ldoceonline.com/dictionary>.

Longman Dictionary of Contemporary English. CD-ROM. 4th edition.

Macmillan English Dictionary Online. Accessed on 3 November 2014.
<http://www.macmillandictionary.com>.

Oxford Advanced Learners' Dictionary (OALD) Online. Accessed on 3 November 2014.
<http://www.oxfordlearnersdictionaries.com>.

Bibliografia adicional

AL-AJMI, H. 2008. 'The Effectiveness of Dictionary Examples in Decoding: The Case of Kuwaiti Learners of English'. *Lexikos*, 18: 15–26.

BAISA, V. AND V. SUCHOMEL. 2014. 'SKELL: Web Interface for English Language Learning'. In *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Tribun EU, 63–70.

BÉJOINT, H. 1981. 'The Foreign Student's Use of Monolingual English Dictionaries'. *Applied Linguistics*, 2.3: 207–222.

BOGAARDS, P.; W. van der KLOOT. 2002. 'Verb Constructions in Learners' Dictionaries'. In A. Braasch and C. Povlsen (eds), *Proceedings of the Tenth Euralex International Congress*. Copenhagen: CST, 747–757.

CHAN, A. 2011. 'Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word'. *International Journal of Lexicography*, 25.1: 68–94.

DAVIES, M. 2008. *The Corpus of Contemporary American English (COCA): 410+ Million Words, 1990-Present*. Accessed on 10 June 2012. <http://www.americancorpus.org>.

DZIEMIANKO, A. 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Tübingen: Niemeyer.

_____. 2012. *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*. Poznań: Adam Mickiewicz University Press.

FRANKENBERG-GARCIA, A. 2005. 'A Peek into What Today's Language Learners as Researchers Actually Do'. *International Journal of Lexicography*, 18.3: 335–355.

_____. 2011. 'Beyond L1-L2 Equivalents: Where do Users of English as a Foreign Language Turn for Help?'. *International Journal of Lexicography*, 24.1: 97–123.

_____. 2012. 'Learners' Use of Corpus Examples'. *International Journal of Lexicography*, 25.3: 273–296.

_____. 2014. 'The Use of Corpus Examples for Language Comprehension and Production'. *ReCALL*, 26: 128–146.

HANKS, P. 1987. 'Definitions and Explanations'. In J. Sinclair (ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London and Glasgow: Collins ELT, 116–136.

HERBST, T. AND G. STEIN. 1987. 'Dictionary Using Skills: a plea for a New Orientation in Language Teaching'. In A. Cowie (ed.), *The Dictionary and the Language Learner*. Tübingen: Niemeyer, 115–127.

HUMBLÉ, P. 2001. *Dictionaries and Language Learners*. Frankfurt: Haag & Herchen.

KILGARRIFF, A., P. RYCHLÝ, P. SMRZ AND D. TUGWELL. 2004. 'The Sketch Engine'. *Proceedings of EURALEX*. Lorient, France, 105–116.

LAUFER, B. 1993. 'The Effects of Dictionary Definitions and Examples on the Comprehension of New L2 words'. *Cahiers de Lexicologie*, 63: 131–142.

NESI, H. 1996. 'The Role of Illustrative Examples in Productive Dictionary Use'. *Dictionaries*, 17: 198–206.

SUMMERS, D. 1988. 'The Role of Dictionaries in Language Learning'. In R. Carter and M. McCarthy (eds), *Vocabulary and Language Teaching*. Longman, 111–125.

TUKEY, J. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Apêndice

Folha de referência 4. *As frases abaixo podem ajudá-lo a melhorar suas traduções. Leia as frases 1 para ajudá-lo com a tradução 1, leia as frases 2 para ajudá-lo com a tradução 2, e assim por diante. Por favor, não escreva nada na folha de referência.*

1. It is the first time she's ever been here.

It is the first time we have attended the exhibition.

This is a brand new idea, it is the first time I have presented it in public.

If it is the first time you have had the symptoms, consult your doctor.

2. The tank was about ten metres wide and fifteen metres long.

The wall was roughly 80 metres long and more than 3 metres high.

The Panama Canal lock chambers are 305 metres long and 33.5 metres wide.

The tunnel was about four hundred metres long.

3. Her great-grandmother would not only approve of Julia's plan but applaud it.

She smiled at him, as if she approved of what he said.

He approves of her every action and would do anything for her.

A majority of Americans approve of interracial marriage

4. He has been in charge since April 2000.

They have lived in Edinburgh since 1984.

Since then, I have been writing about and teaching art and theater.

All of the amateur winners have since turned pro.

5. Always trust your instincts.

Don't trust anyone.

Make sure you can trust your friends.

I use their services because I trust their judgment.

6. She gave him a lot of advice and help.

Thank you so much for your words of advice.

The first step is do not be afraid to seek advice.

I was given three pieces of advice: don't write such complicated sentences, slow down and smile.

7. Call him directly at 214-555-049 to give him your response.

Give us a call at 800-989-8255.

For information or directions, call the church at 281-297-5900.

Call me at 719-579-6407 between 9:00 AM and 4:30 PM.

8. Congratulations on your graduation!

First of all, congratulations on the success of your new e-book.

I thanked him for his congratulations on my birthday.

Congratulations on your 50th wedding anniversary.

9. I must telephone the police.

She telephoned the office on a Sunday night using the after-hours number. Bea telephoned the hospital and asked the nurse if Corren had gone home.

The patient telephoned the pharmacy.

10. I tried to vote for you today, but they told us we couldn't vote any more.

I didn't vote for the president's healthcare plan.

Are they voting for the other guy?

Why should people vote for democrats who act like republicans?

Como citar este texto (ABNT):

FRANKENBERG-GARCIA, A. Tradução de Luísa Rabaldo. Dicionários e exemplos de apoio à produção linguística. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 63-86, 2018.

CONSTRUIR PARA DESCONSTRUIR E VOLTAR A CONSTRUIR:
ELABORAÇÃO COLABORATIVA DE
UM DICIONÁRIO ESCOLAR DE CIÊNCIAS¹²

*Build to deconstruct and re-construct: collaborative development
of a school science dictionary*

Rosa Estopà³

Tradução: Tiele Kowarlevski⁴

Supervisão: Líliam Ramos⁵

Revisão técnica: Cleci Bevilacqua⁶

Resumo: Este artigo apresenta a metodologia colaborativa utilizada na construção de um dicionário escolar de ciências para crianças de 5 a 8 anos, baseada no *corpus* de definições feitas por crianças em língua catalã. O trabalho foi conduzido num período de 3 anos com 500 crianças que definiram e descreveram os termos mais básicos da ciência. A partir do *corpus*, selecionamos apenas a informação acumulada, cientificamente positiva, e, assim, elaboramos as definições sem acrescentar informações novas. Finalmente, o artigo mostra a adequação do dicionário através dos resultados de uma prova de compreensão, realizada com 160 estudantes. Estes resultados permitem dar seguimento a projetos que trabalhem a linguagem juntamente com o conhecimento científico desde: não há ciência sem palavras que a representem ou que a comuniquem.

¹ Este artigo foi escrito no desenvolvimento do projeto «Jugando a definir la ciencia», financiado pela Fundação Espanhola para a Ciência e a Tecnologia, Programa de Cultura Científica e da Inovação, 2011, FECYT-FCT-11-2501; IP: Rosa Estopà: «Jugando a definir la ciencia (2): Laboratorios de palabras» (Fundación Española para la Ciencia y la Tecnología, Programa de Cultura científica y de la Innovación 2012, FECYT-FCT-12-4482; IP: Rosa Estopà); e pelo Departament d'Economia i Coneixement, Secretaria d'Universitats i Recerca, 2012, 2012ACDS00050, IP: Rosa Estopà: «Jugant amb les paraules de la ciència». Para mais informações, é possível consultar a página do projeto: <http://defciencia.iula.upf.edu>. Agradeço sinceramente a Segundo Moyano e Iria da Cunha pela leitura do artigo.

² A presente tradução foi autorizada para ser publicada em português pela autora Rosa Estopà. Referência bibliográfica completa do artigo original: ESTOPÀ, R., Construir para deconstruir y volver a construir: elaboración colaborativa. In: *Enseñanza de las ciencias*, Núm. 32.3 (1) (2014), p.263-282. Disponível em: <https://ensciencias.uab.es/ojs-ensciencias/ensciencias/article/view/1344>. Último acesso em 27 jan. de 2019.

³ Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona rosa.estopa@upf.edu.

⁴ Bacharel em Letras – Português/Espanhol, UFRGS.

⁵ Professora do Departamento de Línguas Modernas e do do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁶ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

Palavras-chave: dicionários escolares, *corpus* escolar, terminologia escolar, ciência, língua.

Abstract: This paper presents the collaborative methodology used in the creation of a school dictionary of science for children aged between 5 and 8 years, based on a corpus of definitions given by children in the Catalan language. The work was conducted over a period of 3 years with 500 children who defined and described the most basic terms of science. From the corpus, we selected only the accumulated information which was scientifically positive and thus produced definitions without adding new information. Finally, the article demonstrates the suitability of the dictionary through the results of a lexical comprehension test carried out with 160 students. These results allow the carrying out of future plans which work the language with scientific knowledge from the beginnings: there is not science without words to represent it or to communicate it.

Key words: Dictionaries for *Children*, Children's Corpus, Academic Terminology, Science, Language.

1. Introdução

A ciência é representada, transferida, explicada e aprendida por meio de palavras especializadas, também conhecidas como termos. O acesso ao conhecimento especializado permite o uso adequado e preciso da terminologia (ESTOPÀ; VALERO, 1999). O processo de especialização, no entanto, é progressivo e tem início nos primeiros anos de vida de uma pessoa (VIGOTSKI, 1962; CARRETERO, 1993; GOLOMBEK, 2006). Assim, na medida em que o conhecimento sobre um conceito vai aumentando, o significado do termo que o designa também vai evoluindo. Izquierdo (2001) destaca que “pretender fixar de uma vez por todas um significado é suprimir seu dinamismo, sua capacidade de transformação e de adaptação e é convertê-lo em um fóssil cognitivo que já não serve para pensar”. Por isso, trabalhar a linguagem conjuntamente com o conhecimento científico desde o início, é crucial para qualquer futura profissão científica, já que não há ciência sem palavras para referir-se a ela e tampouco sem palavras para comunicá-la (GUTIÉRREZ, 1998).

De fato, nas aulas em que se transmite a ciência, tanto os professores quanto os alunos costumam passar muito tempo escutando, lendo e escrevendo sobre ela (LEMKE, 1997; WATT, 1989). As mensagens orais e escritas refletem formas científicas de aproximar-se de objetos, fatos, ações e teorias científicas. Izquierdo e Aliberas (2004, p. 117) alertam sobre a existência de graves deficiências comunicativas quando as crianças falam de ciência. E que ainda é frequente a simples resposta de que “os problemas de língua devem ser resolvidos pelo professor de língua”. Afirmação a qual nos opomos, já que o trabalho holístico na escola potencializa a interdependência natural que existe entre as palavras (língua) e o conhecimento (ciência).

Muitas vezes as crianças, também as muito pequenas, perguntam (e se perguntam) sobre o significado de palavras especializadas: O que é uma *cratera*?, Para que serve um *microscópio*?, De que cor é o nosso *cérebro*?, O *universo* e o *espaço* são a mesma

coisa?, O que significa *neurônio*?... Trabalhar o vocabulário especializado é uma tarefa prioritária desde as primeiras etapas do ensino (MALLART, 2002; ILARI, 2008). Todo o professor que se disponha a ajudar o aluno a ser mais competente no conhecimento e na prática de uma língua, valoriza a importância de usar o dicionário em aula – embora logo possa surgir o dilema sobre qual obra de referência deverá utilizar e, inclusive, se para isso deve ser utilizado o dicionário tradicional. No entanto, nem sempre este axioma pode se sobrepor ao ensino do conhecimento científico. Contudo, na explicação da matemática, do meio, da história, etc., o dicionário – e o trabalho com o vocabulário – pode ter a mesma relevância que tem na aula de língua:

Given the importance background knowledge and the fact that vocabulary is such an essential aspect of it, one of the most crucial services that teachers can provide, particularly for students who do not come from academically advantage backgrounds, is systematic instruction in important academic terms. (MARZANO; PICKERING, 2005, p. 3).

Segundo esses autores, “without a basic knowledge of these terms, students will have difficulty understanding information they read or hear. Knowledge of important terms is critical to understanding any subject” (MARZANO; PICKERING, 2005, p. 2-3). Outros autores como Locke (1992), Lemke (1997), Izquierdo e Aliberas (2004) também defendem que as dificuldades linguísticas produzidas, quando se fala de ciência em aula, devem ter suas soluções discutidas de forma explícita.

Nesse sentido, Carsleton e Marzano (2010, p. 1) estabeleceram uma escala de seis etapas para trabalhar o vocabulário acadêmico em sala de aula, as quais nos referimos com o termo *escala Tennessee* (ESTOPÀ, 2013). O primeiro nível da *escala* consiste em proporcionar uma descrição, uma explicação ou um exemplo de cada termo novo. Para esse primeiro passo, o dicionário – que é o livro que explica o que significam as palavras de uma língua ou de um determinado tema – deveria ser um dos recursos básicos de consulta em sala de aula.

Não obstante, muitas línguas – como é o caso do castelhano, do catalão, do galego e do basco – não dispõem de dicionários para todos os níveis escolares. No âmbito da lexicografia escolar em catalão – e também nas demais línguas mencionadas –, existe uma ausência de dicionários para as crianças que estão iniciando a alfabetização (ou dos chamados, por especialistas como Prado (2000), *dicionários infantis*). De fato, nas escolas do Estado espanhol, o dicionário não é introduzido no currículo até o terceiro ano e é trabalhado apenas no quarto e quinto ano do ensino fundamental. No entanto, isso não

acontece em todos os sistemas escolares⁷. Como exemplo, destacamos a política pedagógica atual do Brasil.

Em 2006, no Programa Nacional do Livro Didático (PNLD), o governo do Brasil incluiu, de maneira prioritária, uma comissão para a avaliação e selecção de dicionários escolares⁸. Essa comissão distinguiu três tipos de dicionários gerais escolares partindo do critério do público alvo (CARVALHO; BAGNO, 2011; KRIEGER, 2012):

- Dicionários de tipo I: alunos do primeiro e segundo ano do ensino fundamental. São dicionários adequados para a fase inicial da alfabetização.
- Dicionários de tipo II: alunos do terceiro e quarto ano do ensino fundamental. São dicionários úteis para a fase de consolidação do domínio da escrita.
- Dicionário de tipo III: alunos de quinto a oitavo ano, que correspondem aos últimos anos do ensino fundamental⁹.

Com relação aos números de entradas, ficou estabelecido que os dicionários de tipo I teriam entre 1000 e 3000 entradas; os dicionários de tipo II teriam entre 3.500 e 10.000 entradas; e os de tipo III entre 19.000 e 35.000. Essa comissão propôs, também, uma série de critérios (quanto à letra, à inclusão de ilustrações, à redação da definição, etc.) que deveriam ser cumpridos por cada tipologia para corresponder às necessidades pedagógicas (CARVALHO, 2011). Por exemplo, foi possível chegar à conclusão de que as nomenclaturas dos dicionários de tipo I ou II não podiam ser versões reduzidas de dicionários gerais para adultos ou de dicionários de tipo III. A microestrutura dos dicionários tipo I devia ser muito didática, de modo que o aluno entendesse a definição e, portanto, as entradas deveriam ser descritas com palavras simples. A comissão, além disso, recomendou o uso de definições oracionais¹⁰ para selecionar os dicionários adequados para esses anos iniciais (COWIE, 2000).

No âmbito da lexicografia escolar em catalão, só existem dicionários escolares gerais de língua destinados a crianças a partir de 9 anos, ou seja, são de tipo II e, sobretudo,

⁷ No âmbito anglo-saxão ou francófono, citando dois exemplos próximos, as crianças contam com um grande repertório de dicionários adequados para diversas idades. Os títulos são reveladores: *Oxford Very First Dictionary*; *Oxford First Dictionary*; *Mon premier Larousse: le dictionnaire de 4-7 ans*; *Mon premier dictionnaire: le Robert Benjamin 6-8 ans*; etc.; e também especializados: *S Maths Dictionary for kids* (EATHER, 2013) ou *Science Dictionary for kids* (WESTPHAL, 2009).

⁸ De fato, a inclusão de dicionários para o ensino fundamental nas escolas públicas nacionais do Brasil aconteceu em 2001 por meio de minidicionários de uso escolar, mas foi somente após o PNLD de 2006, que houve um avanço significativo com a criação de uma comissão específica, que analisou e selecionou dicionários e estabeleceu diretrizes para selecionar ou elaborar dicionários. Essas diretrizes podem ser consultadas na seguinte página: <http://portal.mec.gov.br>

⁹ Em 2012, essa mesma comissão ampliou a tipologia de dicionários didáticos de língua portuguesa e estabeleceu um dicionário tipo IV para os alunos do ensino médio.

¹⁰ Em um momento em que a maioria dos dicionários analisados por esta comissão apresentava um padrão de definição clássico, estabelecendo uma equivalência entre a definição e a entrada que reflete o *Princípio da substituição* (LANDAU, 1984; WEINREICH, 1984).

de tipo III. Os dicionários mais significativos da lexicografia escolar catalã recente são os sete listados abaixo:

BARCANOVA (2009). *Diccionari Barcanova de la llengua escolar*. Segundo seu prólogo está destinado a alunos do ensino fundamental e contém 20.000 entradas. Seria um dicionário de tipo III.

CRUÏLLA (2004). *Diccionari bàsic de la llengua catalana*. Os destinatários são alunos dos primeiros anos do ensino fundamental e dos primeiros anos do ESO – últimos 4 anos do ensino fundamental. Também oferece em torno de 20.000 entradas. Seria um dicionário de tipo III.

ENCICLOPÈDIA CATALANA (1995) DIDAC. *Diccionari de català*. Em seu prólogo, não está informada a idade das crianças às quais se dirige o dicionário, só especifica que é destinado ao ensino fundamental. Apesar disso, o fato de conter 15.282 entradas, permite-nos dizer que é destinado aos últimos anos do ensino fundamental. Seria, então, um dicionário de língua geral situado entre os tipos II e III.

EUMO Editorial (1990). *Primer diccionari*. É recomendado para crianças de 8 a 12 anos. Contém 9.246 entradas. Seria um dicionário de língua geral de tipo II.

ONDA (2004) *Diccionari junior nou*. Oferece 26.000 entradas e é destinado aos primeiros anos do ensino fundamental e ao primeiro ciclo da ESO (6º ano). É um dicionário de tipo III.

SANTILLANA Grup Promotor. (2012) *Nou Diccionari escolar de la llengua catalana*. Contém 24.000 entradas e não especifica para que idades é adequado. É, portanto, um dicionário de tipo III.

VOX (2009). *Diccionari escolar de la llengua catalana*. Contém 19.000 entradas. Também é um dicionário de tipo III.

Por isso, todos os dicionários escolares em língua catalã, seguindo o critério do número de entradas, são do tipo III, exceto um – o *Primer diccionari* de EUMO – que seria de tipo II. Além disso, todos são de língua geral, dado que a maior parte da terminologia acadêmica que incorporam encontra-se no mesmo nível que o léxico geral da língua, sem nenhuma marca temática orientadora.

Na lexicografia espanhola o panorama não é muito diferente. Castillo e García Platero (2003) certificam três obras lexicográficas de tipo I, mas ressaltam que “não são dicionários propriamente ditos, já que as palavras não têm definição, pois são as ilustrações contidas nelas que remetem aos distintos vocábulos”. Trata-se, então, de dicionários em imagens, como o *Imaginario. Diccionario en imágenes para niños* (SM, 1992), *Mis primeras 1000 palabras* de Ferrándiz (Libsa, 1998) ou o *Pictoriccionario. Diccionario en imágenes* (Santillana, 1995). Este último com algumas palavras que apresentam uma definição básica.

Em relação aos dicionários de ciências – ou de matemática –, em catalão o vazio é muito maior, já que não conhecemos a existência de nenhum dicionário com essas

características para alunos de ensino fundamental e, menos ainda, especificamente, para alunos dos primeiros anos desse nível de ensino. Referenciamos um vocabulário de matemática *on-line* (<http://www.xtec.cat/~sgracia5/>), elaborado por Sergio Gracia, que faz parte de seu trabalho de conclusão de curso, para realizar um plano de acolhida a estudantes estrangeiros. Esse vocabulário baseia-se na lista de palavras que acompanha o currículo oficial do ensino fundamental (na área de matemática) do Reino Unido. Esse recurso, que tem alguns problemas lexicográficos e operacionais, oferece uma primeira aproximação do vocabulário que se usa nas aulas em três níveis, representados por três cores diferentes. A língua espanhola também carece desse tipo de obra. Não encontramos dicionários atuais de ciências, elaborados na Península Ibérica, que sejam independentes e cujo público destinatário seja das primeiras etapas do ensino fundamental. No *Instituto de la Sordera* no Chile, pelo contrário, foi elaborada uma série de dicionários escolares visuais e eletrônicos sobre diferentes âmbitos científicos para crianças (<http://www.institutodelasordera.cl/enlaces/webh/>). Apesar dos problemas técnicos de acessibilidade do portal, essa experiência representa um primeiro passo.

Em outras práticas escolares, em contrapartida, existem projetos lexicográficos em papel, mas também recursos em suporte digital acessíveis pela internet. Para citar apenas algumas referências, destacamos *Science Dictionary for kids* (de 4 a 9 anos) de Laurie E. Weestphal (2009), *Scholastic Science Dictionary* de Burger e Brunner (2000), *Math Dictionary for kids* de Fitzgeralds (2011), *First Illustrated Science Dictionary* de Khan (2013), *Junior Science Dictionary (Maths and Science Dictionaries)* de Glover (2011). E, em especial, cabe destacar o trabalho do editorial Oxford que, nos últimos anos, publicou diferentes dicionários escolares classificados segundo a disciplina e a idade do aluno. Para as crianças de 5 a 8 anos, por exemplo: *Oxford First Illustrated Science Dictionary* (2001), *Oxford First Illustrated Maths Dictionary* (2001), *Oxford First Science Dictionary* (2001), *Oxford First Maths Dictionary* (2001). Para as crianças de 9 a 11 anos: *Oxford Primary Illustrated Science Dictionary* (2003), *Oxford First Primary Maths Dictionary* (2003); e referente ao ensino médio, *Oxford Student's Science Dictionary* (2006)¹¹.

O *Oxford First Illustrated Science Dictionary*, por exemplo, é um dicionário de 200 palavras básicas da ciência com definições e ilustrações para crianças de 4 a 6 anos, tal como pode-se observar na imagem abaixo:

¹¹ Existem, também, diversos portais que oferecem dicionários ou vocabulários de ciências em inglês como, por exemplo, o *Math Dictionary for Kids* de Eather (<http://www.amathsdictionaryforkids.com/>) ou o *Illustrated Mathematics Dictionary* de Pierce, que faz parte de um portal para aprender matemática *Maths is fun* (<http://www.mathsisfun.com/aboutmathsisfun.html>).

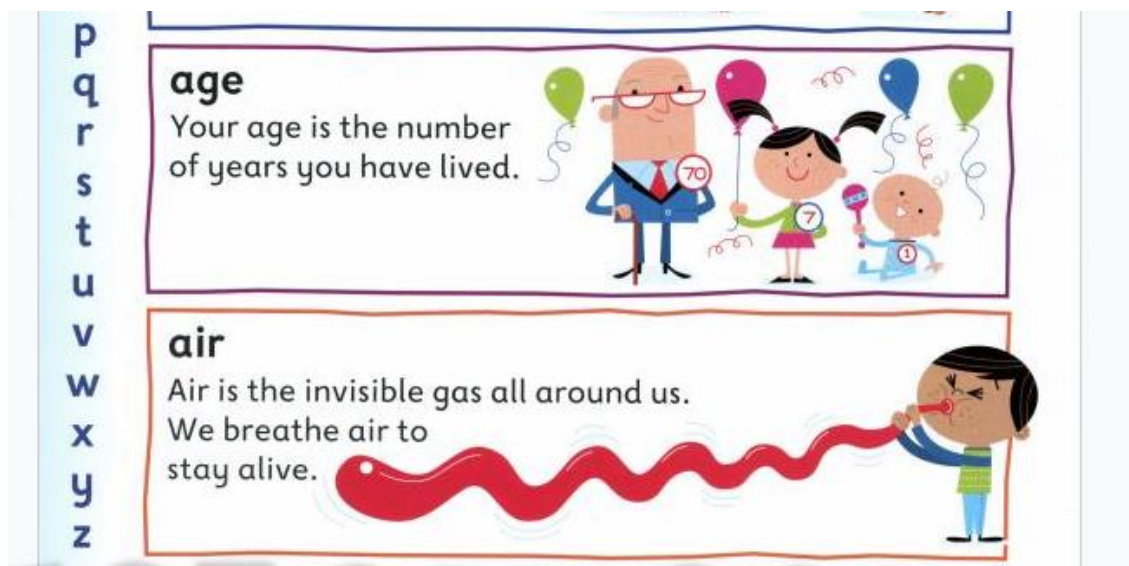


Imagem 1 – Entradas age e air do Oxford First Illustrated Science Dictionary.

Consequentemente, é um axioma lexicográfico dizer que nem todos os dicionários escolares são úteis para todos os usuários, da mesma maneira que é falso pensar que ter um bom dicionário escolar é o suficiente para acompanhar toda a escolaridade obrigatória. As crianças precisam de um dicionário para cada idade ou para cada etapa escolar, já que vão ocorrendo mudanças tanto em sua capacidade cognitiva, quanto em sua compreensão. Sendo assim, não é indicado que se use o mesmo dicionário quando as crianças têm 7, 9 ou 11 anos. E ainda que seja certo que os dicionários escolares têm evoluído muito nos últimos vinte anos (PÉREZ LAGOS, 1998), não temos em nossa tradição lexicográfica uma variedade de dicionários adequados ao ensino fundamental e referentes a âmbitos especializados.

2. Objetivos

O objetivo deste artigo é apresentar a metodologia que utilizamos para a elaboração de um dicionário escolar de ciências destinado a crianças nos anos iniciais do ensino fundamental.

Nossa finalidade foi elaborar um primeiro dicionário de palavras de ciências adequado para os anos iniciais do ensino fundamental destinada à alfabetização. Efetivamente, nos propusemos a construir um dicionário de tipo I com os 100 termos mais gerais da ciência, útil para os anos iniciais do ensino fundamental; ou seja, para crianças de 5 a 8 anos. A criação desse dicionário estava plenamente justificada pelo vazio lexicográfico existente, tanto no que concerne à temática especializada, como aos destinatários da obra.

3. Hipótese

Em nosso projeto, partimos da ideia de que deveríamos planejar um dicionário de termos básicos de ciências para crianças de 6 a 9 anos, que respeitasse o princípio de

adequação da metodologia terminográfica dentro do âmbito da *Teoria Comunicativa da Terminologia* (TCT, CABRÉ, 1999). Segundo esse princípio, cada trabalho terminográfico concreto adota uma estratégia metodológica individualizada em função de suas necessidades terminológicas próprias relacionadas à atividade dos usuários, ao contexto, à temática, aos objetivos e aos recursos disponíveis.

Um dos principais problemas enfrentados pelas crianças é o significado dos termos usados na fala dos professores, nos livros, nas provas e nos exercícios de questões científicas em seu sentido amplo. Para compreender os textos – orais ou escritos – sobre ciências, elas devem entender com suas próprias palavras (por meio de seus esquemas cognitivos) o que significam as palavras relativas às ciências, a que se referem e como são usadas. Consequentemente, nossa hipótese de trabalho foi a seguinte: pode-se conseguir a máxima adequação cognitiva e expressiva de um dicionário com seu público final a partir do conhecimento acumulativo positivo de um coletivo que represente o conjunto de usuários que o utilizarão. Isso significava partir de um *corpus* de representação do conhecimento das crianças. Elas expressam o que sabem, basicamente, com palavras e com desenhos. Por esse motivo, criamos um *corpus* de definições e desenhos de crianças, a partir do qual elaborariamos o dicionário. Planejamos, assim, utilizar uma metodologia colaborativa e acumulativa que partisse do conhecimento científico das crianças.

4. Apresentação do problema: o princípio da adequação

O princípio de adequação é um dos princípios sustentadores da metodologia de trabalho da lexicografia especializada ou da terminologia no âmbito da TCT de María Teresa Cabré (1999). Segundo essa autora, as aplicações terminológicas – tal como são os dicionários especializados – devem adequar-se às necessidades e ao contexto social e linguístico das pessoas a que se destinam. Nas boas práticas terminográficas, a adequação do trabalho terminológico às condições específicas de cada situação é indispensável para assegurar o êxito de um dicionário; entendendo êxito, nesse caso, como a possibilidade de seu uso como ferramenta habitual de consulta.

Os produtos terminográficos ou de lexicografia especializada podem ter características muito variadas e, assim, servir a distintas finalidades. Produzir um glossário, comenta Cabré, pressupõe analisar previamente a situação e as características do projeto que será produzido e, a seguir, esboçá-lo e planejá-lo de acordo com essas características. Cabré pontua que essa maneira de atuar se traduz na aplicação do princípio de adequação. Segundo tal princípio, o trabalho terminológico aplicado não é monolítico, mas variado, em função das circunstâncias temáticas, sociolinguísticas, funcionais e contextuais nas quais será desenvolvido (CABRÉ, 2009).

No campo da lexicografia geral, autores como Bergenholtz e Tarp (1990, 2003), por intermédio de sua teoria moderna das funções lexicográficas, ou Alvar (1982), já haviam reivindicado que os dicionários deveriam adequar-se mais às diversas necessidades que contemplam os usuários:

As editoras não planejaram com rigor uma programação de publicação de dicionários que cubram o maior número possível de necessidades, ficando em aberto, ou com poucos elementos de escolha, áreas que deveriam ser melhor atendidas. Apesar de não parecer, o mercado do dicionário tem baixa concorrência comercial [...] E é bem sabido que a falta de concorrência não melhora a qualidade. (ALVAR, 1982, p. 41).

Dictionaries are considered utility products that are made in order to satisfy certain human needs. Consequently, all theoretical and practical considerations must be based upon a determination of these needs, i.e. what is needed to solve the set of specific problems that pop up for a specific group of users with specific characteristics in specific user situations. (BERGENHOLTZ; TARP, 2003, p. 172).

A ideia de adequação entre dicionário e objetivo de ensino também é um princípio básico da lexicografia pedagógica (HERNÁNDEZ, 1989). Na lexicografia escolar, diversos autores haviam discutido a necessidade de adaptar os dicionários escolares a seus usuários (JAKSON, 2002). Isso implicou uma grande revolução na indústria editorial lexicográfica escolar que, no âmbito da lexicografia catalã, chegou no final da década de oitenta e supôs que os dicionários para a escola não fossem simples reduções dos dicionários de língua geral, assim como havia sinalizado Hernández (1989). No plano internacional, no entanto, foi Sinclair que liderou essa mudança de paradigma com a elaboração do dicionário COBUILD (MOON, 2007), uma proposta de dicionário de aprendizagem que representou uma mudança relevante de metodologia.

No panorama da lexicografia espanhola, na última década do século passado, autores como Hernández (1989) e Martínez Marín (1991) apontavam uma série de defeitos dos dicionários escolares, que Azorín (1998: 78) resume nos seguintes pontos:

- A insuficiente ou nula delimitação, tanto da finalidade didática que acompanha a obra, como do coletivo a que se dirige.
- A inadequada seleção do *corpus* repertoriado.
- A falta de precisão ou excessiva generalização na formulação das definições, que costumam apresentar, algumas vezes, casos de circularidade ou de remissivas.
- Ausência de um critério uniforme na ordem das acepções, assim como na localização dos elementos fraseológicos.
- Graves lacunas na seção relativa à informação gramatical.
- Escassa ou nula presença de exemplos que orientem o uso e favoreçam, portanto, a função codificadora do dicionário.
- Escassa atenção aos aspectos pragmáticos e sociolinguísticos do léxico.

Maldonado (1998, p. 27) completou essa lista referindo-se ao que considera como três pontos fracos: a escassez de obras destinadas aos estudantes, a dificuldade dos docentes para motivar os alunos e a falsa ideia de que todos os dicionários são iguais. Ainda que algumas dessas deficiências se perpetuem em obras contemporâneas, o certo é que hoje em dia dispomos de dicionários escolares muito mais adequados às necessidades das crianças, tanto em relação à macroestrutura (sobretudo na seleção da nomenclatura), como à microestrutura (aspectos gramaticais, definição, exemplos) (CASTILLO; GARCÍA PLATERO, 2003). Atualmente, a maioria dos dicionários escolares *simulam* as explicações do estudante e potencializam a definição oracional. No contexto catalão, por exemplo, os primeiros dicionários elaborados com essa metodologia foram o *Primer Diccionari* (1990), Editora EUMO, organizado por professores prestigiados da Escola de Magistério da Universidad de Vic, e o *Diccionari de català DIDAC* (1995), Enciclopèdia Catalana, sob a direção da lexicógrafa María Teresa Cabré. No entanto, a definição desses tipos de dicionários não deixa de ser uma simulação, já que suas definições *imitam* as narrativas explicativas das crianças, mas não são reais¹².

Diversos trabalhos, entre os quais destacamos os realizados por Pérez Lagos (1998), Ávila Martín (2000), Prado Aragonés (2000) ou os do Grupo de Lexicografia e Linguística de *Corpus* dirigidos por Azorín (1998; 2007), demonstram que, apesar do

indiscutível avanço vivido pela lexicografia escolar, os dicionários – castelhanos e também os catalães – continuam apresentando, no julgamento de seus usuários habituais, deficiências significativas, principalmente quanto à quantidade e qualidade das vozes que registram, ao tratamento da fraseologia, à formulação das definições; à exemplificação e às ilustrações.

Essa inadequação quanto às necessidades do usuário leva a uma baixa utilização do dicionário como ferramenta didática, aspecto destacado pelos mais ilustres teóricos da lexicografia e da didática (REY-DEBOVE, 1971). Em nosso entendimento, o dicionário é introduzido tardiamente nas aulas e as crianças não completam sua conexão com ele. Isso ocorre, em alguns casos, porque não são adequados às necessidades da idade em questão, em outros, porque não se dá a ele a devida importância.

5. Metodologia

Com a finalidade de adequar-se o máximo possível às necessidades de uso das crianças do primeiro e segundo ano do ensino fundamental, nossa estratégia foi trabalhar de maneira colaborativa entre a escola e a universidade para, assim, compilar um *corpus* de definições e desenhos realizados pelas crianças.

¹² Poderíamos estabelecer um paralelismo com a diferença que existe entre um bom exemplo e um contexto de uso real.

A utilização de *corpus* de textos escolares já se mostrou eficiente na lexicografia escolar. Desde 2011, uma equipe do Oxford University Press, em conjunto com o Lexical Computing Department da Oxford University, está realizando experiências muito inovadoras na construção de dicionários para crianças, a partir de *corpus* reais, escritos por elas mesmas. Assim, um *subcorpus* do Oxford Children's Corpus (OCC) – o Beebox – está formado por escritos de cerca de 500 palavras, elaboradas por crianças, distribuídos em três faixas de idade (5 a 8 anos, 9 a 10 anos, mais de 11 anos) sobre um tema livre (WILD *et al.*, 2011 e 2013). A Rádio BBC 2 organiza, anualmente, um concurso de histórias curtas, escritas por crianças com idade entre 5 e 13 anos. O *subcorpus* com textos das crianças procede dos escritos do concurso organizado pela BBC em 2011. De forma que, no início de 2013, o OCC incluía um *subcorpus* de 73.875 histórias (2/12 das histórias correspondiam a crianças entre 5 e 8 anos, 5/12 a crianças entre 9 e 10 anos e os 5/12 restantes a crianças entre 11 e 13 anos) que totalizavam em torno de 32 milhões de ocorrências. Os temas das histórias são muito variados e gerais: pessoas (mãe, avô, irmão, professor etc.), animais, lugares, festas, enredos que dão medo etc. Desde 2006, os lexicógrafos do Oxford University Press utilizam esse tipo de *corpus* para complementar a informação lexicográfica dos seus dicionários para crianças.

Nosso *corpus*, em contrapartida, foi feito com definições-explicações de palavras e desenhos. Para compilar esse material, trabalhamos durante três anos escolares, diretamente na escola: 500 crianças definiram por meio de explicações (breves narrativas) e de desenhos, cerca de 100 termos básicos das ciências. O trabalho de compilação dos dados sempre foi realizado em sala de aula, durante os anos escolares compreendidos entre 2009 e 2012. Participaram três diferentes escolas de Girona, Sabadell e Pineda de Mar. O trabalho foi realizado, principalmente, nas duas primeiras escolas, uma vez que compartilhavam características sociolinguísticas similares; a terceira escola participou somente de cerca de 10% do dicionário, sendo utilizada como escola de controle durante o terceiro ano de trabalho.

Em relação ao nível linguístico dos participantes, foi priorizado o critério de familiaridade e fluência na língua catalã. Por esse motivo, foram selecionadas duas escolas de um contexto social de língua catalã, conhecidas por seu espírito inovador e experimental, tanto nas ciências como na língua. Desses alunos, 85% tem a língua catalã como língua materna, sendo esta a língua habitual de relação entre eles. O segundo critério harmonizador foi o nível social das famílias. São crianças que procedem de famílias com um nível cultural e econômico médio-alto (em torno de 80% dos pais e mães desses alunos têm ensino superior). Optamos por esse tipo de escolas por dois motivos principais: primeiro, porque precisávamos de crianças que, no primeiro ano do ensino fundamental, pudessem se expressar por escrito e, sobretudo, porque, a partir do conhecimento cientificamente adequado e acumulativo dos alunos, teríamos que construir uma definição completa e apropriada que fosse útil ao restante da população da mesma idade. Para tanto, seria mais adequado partir de crianças que, *a priori*, contavam com mais conhecimentos e com um entorno familiar mais letrado, já que se sabe que, nessas idades, as concepções ou

representações estão fortemente influenciadas pelo nível de conhecimento e história contextual (MARÍN, 2003). Sendo assim, não podíamos trabalhar com crianças que elaborassem definições excessivamente simples, do ponto de vista cognitivo, ou ainda lexicalmente pobres, porque, assim, elas não serviriam para todos e, inclusive, necessitaríamos de muito mais *inputs*. Contudo, no terceiro ano do projeto, trabalhamos com uma escola de características opostas, que nos serviu de escola teste; mais especificamente, com uma turma de segundo ano do ensino fundamental, na qual os pais e mães de 35,6% dos alunos eram imigrantes, 22% usavam a língua catalã como língua principal e apenas 2% dos pais tinham ensino superior.

O trabalho foi realizado principalmente no contexto da própria sala de aula dos alunos, com o objetivo de evitar o máximo de interferências externas. No entanto, também foram utilizados outros espaços escolares: a sala de aula de língua, a de projetos e a de artes. Da mesma forma, a maior parte do trabalho foi realizada em um horário periódico e sistematizado, sempre reservado ao dicionário (uma hora semanal).

O leatório foi selecionado após a aplicação de diferentes critérios: a) primeiro foram selecionados os termos mais básicos que constavam em dicionários de tipo II e III e em livros de texto; b) depois, restringiu-se a primeira seleção e comprovou-se que todas as palavras estivessem incluídas na lista de termos de Tennessee, correspondente à idade de 6 a 8 anos (Tennessee Academic Vocabulary del Tennessee Department of Education, 2009); e c) finalmente, a lista foi validada por professores do ensino fundamental de três escolas diferentes. Uma amostra das palavras selecionadas¹³: *agua, aire, balanza, calor, cerebro, experimento, fuego, idea, luna, lupa, memoria, muerte, neurona, ordenador, problema, resta, sonido, termómetro, triángulo, universo*, etc. Algumas são palavras utilizadas na língua comum e fazem parte do vocabulário base.

O resultado foi um *corpus* de desenhos e narrações explicativas reais, de crianças do primeiro e segundo ano do ensino fundamental (5 a 9 anos): o IULATERM Children's Corpus (ICC). O ICC é constituído por 3.550 definições de palavras concretas relacionadas com a ciência, que correspondem a 68.639 ocorrências, e por 3.575 desenhos que ilustram esses conceitos científicos. As definições são breves narrativas que explicam o significado dessas palavras. As crianças trabalharam em papel e toda a informação foi digitalizada, de modo que, agora, está acessível através do portal aberto *El Microscopio*, no seguinte site: <http://defciencia.iula.upf.edu/Microscopi> (ESTOPÀ; CORNUDELLA, 2013)¹⁴.

O ICC foi utilizado na elaboração de cada uma das definições do dicionário. Para construir a definição/explicação final, foi utilizada uma estratégia de acumulação de informação positiva. Assim, foi selecionada toda a informação positiva do ponto de vista científico das definições do ICC, respeitando as palavras usadas pelos estudantes em seus escritos. Para isso, foi excluída qualquer informação *tóxica*, imaginária ou pertencente a

¹³ N.T.: água, ar, balança, calor, cérebro, experimento, fogo, ideia, lua, lupa, memória, morte, neurônio, computador, problema, diferença, som, termômetro, triângulo, universo etc.

¹⁴ Apenas foram corrigidos os erros de ortografia e de gramática.

um conhecimento cultural ou contextual. Na última etapa, foi realizado o controle de conhecimento por especialistas (biólogos, médicos, físicos, engenheiros). Dessa forma, diversos especialistas asseguraram a autenticidade da informação selecionada do ponto de vista científico.

As definições foram elaboradas seguindo o padrão mais recorrente usados pelos estudantes. Em todas as definições foram incluídos exemplos e contextualizações que as crianças haviam mencionado. Esse tipo de informação contribui enormemente para a compreensão do termo. Os exemplos nunca foram inventados ou manipulados; são os mesmos propostos pelas crianças em suas narrações. Na palavra *huevo*¹⁵, por exemplo, está dito:

[...] Los huevos pueden ser de diferente tamaño, por ejemplo, los del avestruz son más grandes que los de la gallina... Nosotros solemos comer huevos de gallina y usamos este alimento para hacer, por ejemplo, una tortilla, un huevo frito o un huevo duro o también cuando queremos hacer una torta, un pastel o magdalenas¹⁶.

E na palavra *cerebro* está dito:

El cerebro es una parte del cuerpo humano que está dentro del cráneo y que sirve para pensar y también para estudiar, leer, escuchar, hablar, andar, trabajar, elegir [...] ¹⁷.

Em relação à inclusão de acepções, foi utilizado o mesmo critério: incluir somente as acepções que as crianças detectaram. No momento em que a criança (ainda que tenha sido uma só) definia uma palavra com uma acepção nova, incluimos no dicionário. É o caso, por exemplo, de *acuario*, *termómetro*, *tierra*, *problema* ou *planta*¹⁸. Por outro lado, em outros casos as crianças concordavam em destacar um significado único, apesar de ser um termo polissêmico. Então, guardávamos somente o significado assegurado pelas crianças, como aconteceu, por exemplo, com *balanza*, *memoria* ou *espacio*¹⁹.

Finalmente, também se considerou que os desenhos são um complemento muito potente das narrações breves, ou seja, das definições. Por um lado, permitem integrar a todas as crianças, pois, caso alguma não soubesse se expressar escrevendo, podia participar

¹⁵ N.T.: ovo.

¹⁶ N.T.: Os ovos podem ser de diferentes tamanhos, por exemplo, os de avestruz são maiores que os de galinha... Nós costumamos comer ovos de galinha e usamos este alimento para fazer, por exemplo, uma tortilla, um ovo frito ou um ovo cozido ou também quando queremos fazer uma torta, um bolo ou cupcakes.

¹⁷ N.T.: O cérebro é uma parte do corpo humano que está dentro do crânio e que serve para pensar e também para estudar, ler, escutar, falar, andar, trabalhar, escolher...

¹⁸ N.T.: aquário, termômetro, terra, problema ou planta.

¹⁹ N.T.: balança, memória ou espaço.

apenas desenhando (o que ocorreu com 4% dos alunos de primeiro ano do ensino fundamental). Por outro lado, os desenhos das crianças mostram com detalhe certos matizes que não ficam refletidos nas explicações. Havia termos extremamente difíceis de serem representados: conceitos abstratos ou intangíveis sem a ajuda de algum instrumento. É complexo desenhar termos como *fuerza*, *aire* ou *neurona*²⁰; em contrapartida, isso não representou um obstáculo. Suas representações refletem a noção comum ocultada por esses termos em níveis iniciais²¹.

6. Resultados: *mi primer diccionario de ciencia*, um dicionário de ciências escolar colaborativo

O *Petit diccionari de ciència* [Meu primeiro dicionário de ciências]²² (PDC) é, então, um dicionário de tipo I especializado, que contém 100 termos básicos, que correspondem a 126 acepções distintas; 13 delas são unidades verbais e o restante são nomes. É destinado a acompanhar a aprendizagem de crianças de primeiro, segundo (e em alguns casos de terceiro) ano do ensino fundamental. Cada uma das entradas inclui a categoria gramatical, uma explicação e um desenho.

A explicação do termo é uma narração sobre como a criança define essa palavra, como a entende, o que significa para ela, como a usa. É uma definição lexicograficamente pouco canônica; posto que sempre é multioracional, situacional e pragmática. A informação está ordenada seguindo a importância que as crianças deram às características semânticas em suas definições. As definições do PDC, muitas vezes, seguem padrões como “X é um” (normalmente seguido de um hiperônimo e nem sempre do mesmo nível) ou “Quando...,” (normalmente para explicar o significado de um verbo). Vejamos alguns exemplos de como são iniciadas as definições no dicionário final:

aire

El aire es transparente e incoloro...²³

brújula

Una brújula es un objeto parecido a un reloj...²⁴

científico científica

²⁰ N.T.: força, ar ou neurônio.

²¹ Por questões de edição, selecionamos apenas um único desenho para o dicionário. Contudo, é possível consultar uma seleção de desenhos por meio de *El Microscopio* ou no site do projeto: http://defciencia.iula.upf.edu/index_eng.htm

²² Publicado também em castelhano por Publicacions de l'Abadia de Montserrat (2013): *Mi primer diccionario de ciencia*. O dicionário está em edição para ser publicado no Brasil como dicionário trilingue (português, espanhol e inglês).

²³ N.T.: O ar é transparente e incoloro...

²⁴ N.T.: Uma bússola é um objeto parecido com um relógio...

Un científico o una científica es una persona...²⁵

contar

Cuando contamos decimos cuántas cosas hay...²⁶

hielo

El hielo es agua congelada...²⁷

número

Un número se usa para contar...²⁸

tiempo

El tiempo es lo que pasa entre dos hechos o también lo que duran las cosas...²⁹

As relações de função ou de parte/tudo são as mais recorrentes entre suas narrações. Como já se sabe, para as crianças interessa mais saber para o que servem as coisas do que o que são, de maneira que a relação funcional é mais relevante que a ontológica. As explicações das crianças, além disso, estão estreitamente relacionadas com o seu contexto experimental. Por isso, é tão importante que as definições recorram aos exemplos e às situações pragmático-comunicativas. Assim, optamos por incluir o maior número de contextos que se aproximasse da criança em sua realidade cotidiana. Essas decisões metodológicas estão respaldadas por especialistas em didática da ciência, como Driver (1983), De Posada (1996) ou Marín (2003), que sustentam que “no sujeito da etapa infantil são mais relevantes as interações que dão lugar à uma estrutura de caráter procedimental e implicitamente muito ligada ao efetivo” (MARÍN, 2003, p. 80).

Reproduzimos abaixo a entrada do dicionário: *estrella*³⁰. Ao analisá-la, constatamos que, na verdade, não está dito o que é uma *estrella* (que é a essência da definição clássica) e sim, como é, onde se pode ver, do que se constitui, quantas existem, exemplos de estrelas. Além disso, nesta definição é muito interessante o detalhe discursivo “...embora não pareça uma estrela como as outras”, porque reflete uma cumplicidade cognitiva com a criança, dado que a muitas crianças surpreende que o Sol seja uma estrela, pois a vemos de dia e não à noite, parecendo muito maior aos nossos olhos, mais brilhante e mais potente que as demais que costumamos contemplar à noite.

²⁵ N.T.: Um cientista ou uma cientista é uma pessoa...

²⁶ N.T.: Quando contamos dizemos quantas coisas tem...

²⁷ N.T.: O gelo é água congelada...

²⁸ N.T.: Um número se usa para contar...

²⁹ N.T.: O tempo é o que passa entre dois acontecimentos ou também o quanto duram as coisas...

³⁰ N.T.: estrela.

estrella

nombre femenino

Una estrella es luminosa. Las estrellas las vemos brillar por la noche en el cielo, como la Estrella Polar. En el espacio hay estrellas de muchos y diferentes tamaños. Las estrellas están formadas de gas. Para nosotros, la estrella más importante es el Sol, aunque no parezca una estrella como las otras. En el Universo, en cambio, hay millones de estrellas.



43

Fiona

Imagem 2 – estrella de Mi primer diccionario de ciencia³¹.

7. Avaliações e valoração do resultado

Sabemos, pelo contato realizado com as crianças que têm utilizado o dicionário, que elas o entendem, gostam dele e que os desenhos as fascinam. Diversos professores também têm comentado que o dicionário parece muito adequado e útil por não haver uma obra parecida em catalão ou castelhano. Contudo, realizamos uma prova experimental para ratificar essas primeiras impressões que, do contrário, corriam o risco de ser excessivamente subjetivas. Com esse objetivo, criamos uma prova de compreensão: queríamos comprovar se, para as crianças de 6 a 8 anos, as explicações elaboradas para o *Mi primer diccionario de ciencia* são mais adequadas que as definições que incluem outros dicionários escolares existentes na lexicografia catalã.

O grupo que realizou a prova foi constituído por crianças do primeiro ano (70 alunos, 3 turmas), segundo ano (91 alunos, 4 turmas) e do terceiro ano (25 alunos, 1 turma) do ensino fundamental de três escolas de Pineda de Mar, Banyoles e Santa Susanna, respectivamente; eram crianças que não tinham participado da elaboração do dicionário. O nível socioeconômico e sociolinguístico dos alunos dessas escolas é variado e acreditamos que pode ser mais ou menos significativo para uma porcentagem elevada de crianças. A escola de Pineda de Mar, como referimos anteriormente, apresenta uma alta proporção de filhos de pais estrangeiros que não têm o catalão como língua materna e com um nível socioeconômico e sociocultural baixo. A escola de Banyoles possui 25% de

³¹ N.T.: Imagem 2. Entrada *estrella* do *Mi primer diccionario de ciencia*

imigrantes, a maioria dos pais é falante de catalão, tem uma situação socioeconômica média e 40% desses pais possui estudo superior. A escola de Santa Susanna tem uma porcentagem de imigração baixíssima (5%), 60% dos pais não tem o catalão como língua materna, seu nível socioeconômico é médio/alto e o nível sociocultural também é médio/alto (70% das crianças tem um dos pais com estudo superior). As três escolas estão situadas em comunidades com muitos serviços e próximas à capital da província.

Selecionamos três termos: *água*, *cérebro* e *telescópio*. Logo, extraímos as definições desses termos de quatro dicionários escolares: EUMO (Tipo II), Santillana (tipo III), DIDAC (tipo III) e *Mi primer diccionario de ciencia* (PDC) (tipo I). Misturamos e oferecemos essas definições em cada pergunta, com uma ordem distinta. Na prova, pedimos aos participantes que pusessem uma cruz verde na explicação que melhor entendessem e uma cruz vermelha na que menos entendessem ou que não entendessem. Também pedimos que, em todas as definições, marcassem em vermelho as palavras que não compreendessem. A prova foi aplicada por seus professores em sua aula habitual. Reproduzimos as perguntas da prova de compreensão:

Què és un cervell? [O que é um cérebro?]

- A. Part més gran de l'encèfal dels vertebrats, situada a la zona del davant i de dalt del crani. [Parte maior do encéfalo dos vertebrados, situada na zona da frente e de cima do crânio.]
- B. El cervell és una part del cos humà que és dins del crani i que serveix per a pensar i també per a estudiar, llegir, escoltar, parlar, caminar, treballar, escollir... El cervell és tou, arrodonit i està format per neurones. Hi ha animals que també tenen cervell, però és diferent del de les persones. Si no tinguéssim cervell, no seríem persones. [O cérebro é uma parte do corpo humano que está dentro do crânio e que serve para pensar e também para estudar, ler, escutar, falar, andar, trabalhar, escolher... O cérebro é mole, arredondado e está formado por neurônios. Há animais que também têm cérebro, mas é diferente do das pessoas. Se não tivéssemos cérebro, não seríamos pessoas.]
- C. El cervell és un òrgan tou, de forma arrodonida i amb molt plects que hi ha dins del cap de les persones. Molts animals també tenen cervell. És una part de l'encèfal. El pensament i els moviments voluntaris depenen del cervell. [O cérebro é um órgão mole, de formato arredondado e com muitas dobras, que está dentro da cabeça das pessoas. Muitos animais têm cérebro. É uma parte do encéfalo. O pensamento e os movimentos voluntários dependem do cérebro.]
- D. Òrgan de la intel·ligència que es troba a l'interior del cap de l'home i dels animals. [Órgão da inteligência, que se encontra no interior da cabeça do homem e dos animais.]

A primeira (A) procede do dicionário Santillana, a segunda (B) do PDC, a terceira (C) do DIDAC e a última (D) do dicionário EUMO.

Què és el *agua*? [O que é a água?]

- A. L'aigua és un líquid incolor que no té gust ni fa olor. L'aigua és formada d'oxigen i hidrogen. És un gran dissolvent. Es torna sòlida a 0 °C i forma el gel, i bull a 100 °C. La seva fórmula és: H₂O. [A água é um líquido incolor, que não tem gosto, nem cheiro. A água é formada por oxigênio e hidrogênio. É um grande solvente. Fica sólida a 0 °C e forma o gelo, e ferve a 100 °C. Sua fórmula é: H₂O.]
- B. Líquid format per hidrogen i oxigen, no té color, olor ni sabor. I es troba a la natura en els rius, mars i llacs. [Líquido formado por hidrogênio e oxigênio, não tem cor, cheiro, nem sabor. E se encontra na natureza e nos rios, mares e lagos.]
- C. Líquid que hi ha als rius, als llacs i al mar: *les aigües del riu baixaven brutes*. [Líquido que tem nos rios, nos lagos e no mar: *As águas do rio desciam sujas*.]
- D. L'aigua és un líquid transparent que podem beure i que també podem fer servir per a rentar-nos o regar les plantes. L'aigua de mar és salada i l'aigua de riu o de pluja, no. Les persones, els animals i les plantes necessitem aigua per a viure. [A água é um líquido transparente que podemos beber e que também podemos usar para nos lavar ou regar as plantas. A água do mar é salgada e a água do rio ou da chuva, não. As pessoas, os animais e as plantas necessitam de água para viver.]

A primeira (A) procede do diccionário DIDAC, la segona (B) del Santillana, la tercera (C) del diccionário EUMO i la última (D) del PDC.

Què és un *telescopi*? [O que é um telescópio?]

- A. Instrument que serveix per observar les estrelles, els planetes i els satèl·lits. [Instrumento que serve para observar as estrelas, os planetas e os satélites.]
- B. Un telescopi és un aparell òptic per a mirar objectes que estan molt lluny, sobretot les estrelles i altres cossos que hi ha al cel. Alguns telescopis són molt grans i estan tancats dins d'un edifici amb una cúpula que s'obre quan es vol mirar el cel. [Um telescópio é um aparelho ótico para olhar objetos que estão muito longe, sobretudo as estrelas e outros corpos que estão no céu. Alguns telescópios são muito grandes e estão encerrados dentro de um edifício com uma cúpula que se abre quando se quer ver o céu.]
- C. Instrument òptic que s'utilitza per observar objectes llunyans, per exemple els estels. [Instrumento ótico que se utiliza para observar objetos distantes, por exemplo as estrelas.]
- D. Un telescopi és un aparell que enfoca el que està molt lluny per veureu-ho millor, com ara la Lluna, les estrelles, les constel·lacions i els planetes. En astronomia s'utilitza el telescopi per a observar i investigar l'Univers. [Um telescópio é um aparelho que enfoca o que está muito longe para vê-lo melhor, como a Lua, as estrelas, as constelações e os planetas. Em Astronomia, se utiliza o telescópio para observar e pesquisar o Universo.]

A primeira (A) procede do dicionário EUMO, a segunda (B) do DIDAC, a terceira (C) do dicionário Santillana e a última (D) do PDC.

Os resultados das provas nos distintos níveis escolares estão reunidos abaixo nas tabelas 1, 2 e 3. Indicamos, em cada caso, o número absoluto de estudantes que selecionaram uma opção e também a porcentagem, para que seja possível comparar as respostas entre as distintas idades, dado que o número de participantes por ano escolar não é idêntico:

1º	Santillana		DIDAC		EUMO		PDC		Total de estudantes
	+	-	+	-	+	-	+	-	
cérebro	1 1,4%	43 61,4%	3 4,3%	21 30%	--	--	66 94,3%	6 8,6%	70 100%
água	3 4,3%	11 15,8%	5 7,1%	58 82,9%	2 2,9%	1 1,5%	60 85,8%	--	70 100%
telescópio	2 2,9%	17 24,3%	4 5,7%	40 57,1%	15 21,5%	9 12,9%	49 64,3%	4 5,7%	70 100%

Tabela 1 – Resultados das provas dos estudantes de três turmas de primeiro ano do ensino fundamental.

2º	Santillana		DIDAC		EUMO		PDC		Total de estudantes
	+	-	+	-	+	-	+	-	
cérebro	3 3,3%	25 27,5%	8 8,8%	54 59,3%	20 22%	5 5,5%	60 65,9%	7 7,7%	91 100%
água	6 6,6%	8 8,8%	1 1,1%	72 79,1%	29 31,9%	5 5,5%	55 60,4%	6 6,6%	91 100%
telescópio	7 7,7%	15 16,5%	7 7,7%	50 54,9%	32 35,2%	14 15,4%	45 49,5%	12 13,2%	91 100%

Tabela 2 – Resultados das provas dos estudantes de quatro turmas do segundo ano de ensino fundamental.

3º	Santillana		DIDAC		EUMO		PDC		Total de estudantes
	+	-	+	-	+	-	+	-	
cérebro	--	19 76%	5 20%	--	--	1 4%	25 100%	--	25 100%
água	3 12%	--	--	22 92%	7 28%	1 4%	15 60%	1 4%	25 100%
telescópio	6 24%	4 16%	2 8%	17 68%	5 20%	3 12%	12 48%	1 4%	25 100%

Tabela 3 – Resultados das provas dos estudantes de uma turma de terceiro ano do ensino fundamental.

Observa-se que, nas três tabelas, as definições entendidas pela maioria dos alunos são as do PDC e as menos entendidas são as do dicionário DIDAC e as do dicionário Santillana. Este resultado não deveria nos surpreender, posto que o DIDAC e o Santillana são os dois dicionários escolares com mais entradas e que representariam dicionários dos tipos II-III e III, respectivamente. No entanto, em contrapartida, são surpreendentes os resultados referentes ao DIDAC porque dos três dicionários é o que, explicitamente, tem a intenção – como dito no prólogo – de elaborar definições oracionais e incluir exemplos próximos às crianças do ensino fundamental. A parte gráfica parece aproximada das crianças de ensino fundamental, das primeiras etapas, mas nem sempre é o caso da informação cognitiva, tampouco para as crianças de terceiro ano. Talvez um dos parâmetros que não tenha se adequado às necessidades das crianças participantes seja alguma terminologia usada nas definições. Por isso, é relevante, e lógico, constatar que as definições que elas recusam porque não entendem são as que incluem mais palavras marcadas em vermelho, ou seja, as que têm mais palavras que desconhecem – por se tratarem de unidades terminológicas – ou as definições que incluem símbolos e fórmulas. Das crianças que participaram da prova, 95% marcaram em vermelho as seguintes palavras:

Cerebro:

Santillana: encèfal [encéfalo] e vertebrat [vertebrado]

EUMO: òrgan [órgão]

DIDAC: encèfal [encéfalo], òrgan [órgão], plecs [dobras], pensament [pensamento]

PDC: neurones [neurônios]

Agua:

Santillana: hidrogen [hidrogênio], oxigen [oxigênio]

EUMO: --

DIDAC: 0 °C, 100 °C, H₂O, sòlida [sólida], formula [fórmula], hidrogen [hidrogênio], dissolvent [solvente]

PDC: --

Telescopio:

Santillana: òptic [ótico], llunyà [distante]

EUMO: satèl·lit [satélite]

DIDAC: òptic [ótico], cúpula [cúpula], cossos [corpos]

PDC: --

Em particular, um aluno sinalizou as seguintes palavras como desconhecidas ou que dificultam sua compreensão:

- *Agua*: incolor (incolor), líquid incolor (líquido incolor) (DIDAC).
- *Telescopio*: astronomia (astronomia), constel·lació (constelação) (PDC).

As palavras *astronomia* e *constelação* que o PDC usa na definição de *telescopio* seguramente são o motivo pelo qual, nessa definição, embora siga sendo a mais compreensível, as crianças reduziram o nível de compreensão (primeiro ano: 64,3%, segundo ano: 49,5%, terceiro ano: 48%) em relação a outras palavras estudadas.

Em contrapartida, a compreensão das definições parece não ter relação nem com o número de palavras usadas, nem com a quantidade de orações que a definição apresenta. Os dados nos mostram que os dicionários que utilizam uma explicação com mais de uma oração – DIDAC e PDC – são, em relação ao grau de compreensão escolar, opostos, tal como ilustrado na tabela 4:

Número de orações				
	Santillana	EUMO	DIDAC	PDC
cérebro	1	1	4	4
água	1	1	5	5
telescópio	1	1	2	2

Tabela 4 – Número de orações aplicadas em cada definição.

8. Considerações finais e linhas de trabalho futuro

A metodologia acumulativa e colaborativa utilizada, baseada em um *corpus* de definições feitas por crianças para elaborar o dicionário, se mostrou adequada: primeiro construir, depois desconstruir e, finalmente, voltar a construir. As crianças elaboram individualmente suas definições, nós desconstruímos todas elas e acumulamos o conhecimento e, por fim, selecionamos a informação cientificamente pertinente. As definições do PDC, um primeiro dicionário de ciência para crianças de 5 a 8 anos, são, então, explicações baseadas no saber acumulativo controlado das crianças dessa idade. O teste experimental ratificou que as crianças dessa idade entendem melhor o PDC que outros dicionários escolares existentes na lexicografia catalã, destinados, desde o nosso ponto de vista, a um nível escolar superior.

A partir dessa experiência, criamos o Club LEXIC (ESTOPÀ e CORNUDELLA, 2013; ESTOPÀ, 2013), que permite alimentar o *corpus* de definições, tanto em relação à idade das crianças, como à quantidade de textos. Esta ferramenta já está à disposição das escolas desde o período letivo 2013-2014 e tem acesso livre no seguinte *site*: <http://defciencia.iula.upf.edu/ClubLexicografs> Trata-se de uma plataforma digital que permite trabalhar em castelhano e catalão, embora seja facilmente adaptável para outras línguas. Devido às leis de proteção e de direitos autorais de menores, foi necessário criar dois perfis de usuários distintos: um para aquelas crianças que colaboraram anônima e pontualmente na elaboração de futuros dicionários colaborativos e outro perfil, com mais funcionalidades, para as crianças que fossem colaborar de forma mais constante, incentivando-as a fazer parte do Club LEXIC. Além disso, ser membro permite poder

dispor de um material diacrônico e trabalhar na escola, posteriormente, com esse dados, como se fosse um portfólio digital, individualizado, sobre léxico.

Referências

ALVAR, M. *Función del diccionario en la enseñanza de la lengua*. Revista de Bachillerato 22, p. 49-53, 1982.

AZORÍN, D. *¿Para qué usan los diccionarios los escolares?: Reflexiones a propósito de una encuesta*. Actas del congreso Tendencias en la lexicografía española actual. El diccionario como objeto de estudio lingüístico y didáctico. Huelva: Universidad de Huelva, p. 75-100, 2000.

AZORÍN, D. *Los diccionarios didácticos del español desde la perspectiva de sus destinatarios*. Estudios de lingüística 14, p. 19-44, 2000.

AZORÍN D. *La investigación sobre el uso del diccionario en el ámbito escolar*. NEGRI, A. y ALVES, I. M. (orgs.). In: *As Ciências do léxico: lexicologia, lexicografia, terminologia*. Campo Grande: Ed. UFMS, 2007.

BARCANOVA. *Diccionari Barcanova de la llengua escolar*. Barcelona: Barcanova, 2009.

BERGENHOLTZ, H.; TARP, S. *Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions*. *Hermes. Journal of Linguistics*, 31, p. 171-196, 2003.

BURGER, M.; BRUNNER, H. *Scholastic Science Dictionary*. USA: Scholastic Inc., 2000

CABRÉ, M. T. *La terminología. Representación y comunicación: Una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, 1999.

_____. *Terminología y buenas prácticas*, Atti Convegno Assiterm 2009, Publifarum 12, 2010. Disponible en: <http://www.publifarum.farum.it/ezone_articles.php?art_id=161>. Accedido en: enero de 2014.

CARLETON, L.; MARZANO, R. *Vocabulary games for the classroom*. Bloomington: Marzano Research Laboratory, 2010.

CARRETERO, M. *Constructivismo y educación*. Buenos Aires: Aique, 1993.

CARVALHO, O. *Dicionários escolares: definição oracional e texto lexicográfico*. CARVALHO, O.; BAGNO, M. (orgs.). *Dicionários escolares: políticas, formas e usos*. São Paulo: Parábola, 2011.

CARVALHO, O.; BAGNO, M. (orgs.). *Dicionários escolares: políticas, formas e usos*. São Paulo: Parábola, 2011.

CASTILLO, A.; García Platero, J. M. *La lexicografía didáctica*. Media, A. *La lexicografía española*. Barcelona: Ariel, p. 333-351, 2003.

COWIE, A. P. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press, 2000.

CRUÏLLA. *Diccionari bàsic de la llengua catalana*. Barcelona: Cruïlla, 2004.

DE POSADA, J. M. *Hacia una teoría de las ideas científicas de los alumnos*: influencia del contexto. *Enseñanza de ciencias*, 14 (3), p. 303-314, 1996.

DE VECCHI, G.; GIORDAN, A. *L'Enseignement scientifique*. Comment faire pour que "ça marche"? Nice: Z'édicions, 1996.

DRIVER, R. *The pupil as scientist*. Milton Keynes: Open University Press, 1983.

EATHER, J. *A Maths Dictionary for kids*. 2013 Disponible en: <<http://www.amathsdictionaryforkids.com/>>.

ENCICLOPÈDIA CATALANA. *DIDAC*: Diccionari de català. Barcelona: Enciclopèdia catalana, 1995.

ESTOPÀ, R. *Jugant a definir la ciència*: un diccionari de mots de ciència fet per i per a nens i nenes. *Terminalia* 4, p. 25-33, 2011. Disponible en: <<https://www.raco.cat/index.php/Terminalia/index>> Accedido en: enero de 2014.

ESTOPÀ, R. (dir.) *Petit diccionari de ciència*. Barcelona: Publicacions de l'Abadia de Montserrat, 2012.

ESTOPÀ, R. (dir.). *Petit imaginari de ciència*. Barcelona: Publicacions de l'Abadia de Montserrat, 2012.

ESTOPÀ, R. (dir.). *Mi primer diccionario de ciencia*. Barcelona: Publicacions de l'Abadia de Montserrat, 2013.

ESTOPÀ, R. *Juguem a definir la ciència*: recursos per treballar el lèxic especialitzat a l'escola. SÀNCHEZ FÈRRIZ, M. À. (cur.) In: *La terminologia en les ciències de la vida, en la química i en el món educatiu*. Barcelona: SCATERM, Memòries de la Societat catalana de terminologia, 4, p. 199-209, 2013. Disponible en: <<https://publicacions.iec.cat/repository/pdf/00000200%5C00000030.pdf>>. Accedido en: enero de 2014.

ESTOPÀ, R.; CORNUDELLA, M. *El CLUB LEXIC y el MICROSCOPIO*: plataformas en línea para construir diccionarios científicos colaborativos en un proyecto universidad-escuela. Girona: Uninvest, 2013.

ESTOPÀ, R.; VALERO, T. *Adquisición del conocimiento especializado y unidades de significación especializada en medicina*. In: Panacea@. *Boletín de Medicina y Traducción* 3, p. 277-292, 2002. Disponible en: <<http://www.medtrad.org/panacea.html>>. Accedido en: enero de 2014.

EUMO. *Primer diccionari*. Vic: Eumo editorial, 1990.

- FITZGERALD, ST. *Math Dictionary for kids*. USA: Prufrock Press, 2011.
- GLOVER, D. *Junior Science Dictionary (Maths and Science Dictionaries)*. Londres: Pearson Education, 2011.
- GOLOMBEK, D. *Prólogo*. CHARPAK, G. Et al. In: *Los niños y la ciencia: La aventura de La mano en la masa*. Buenos Aires: Siglo XXI editores, 2006.
- GUTIÉRREZ, B. *La ciencia empieza en la palabra*. Barcelona: Península, 1998.
- HERNÁNDEZ, H. *Los diccionarios de orientación escolar: Contribución al estudio de la lexicografía española*. Tübingen: Max Niemeyer, 1989.
- ILARI, R. *Introdução ao estudo do léxico: Brincando com as palavras*. Campinas: Contexto, 2008.
- IZQUIERDO, M.; ALIBERAS, J. *Pensar, actuar i parlar a la classe de ciències: Per un ensenyament de les ciències racional i raonable*. Bellaterra: Universitat Autònoma de Barcelona, 2004.
- JACKSON, H. *Lexicography: an introduction*. London: Routledge, 2002.
- KHAN, S. *First Illustrated Science Dictionary*. Oxford: Usborne, 2013.
- KRIEGER, M. G. *Dicionário em sala de aula: guia de estudos e exercícios*. Rio de Janeiro: Lexikon, 2012.
- LANDAU, S. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press, 1984.
- LEMKE, J. *Aprender a hablar ciencia: Lenguaje, aprendizaje y valores*. Barcelona: Paidós, 1997.
- LOCKE, D. *Science as Writing*. New Haven: Yale University Press (versión en español: (1997). *La ciencia como escritura*. Madrid: Cátedra), 1992.
- MALDONADO, C. *El uso del diccionario en el aula*. Madrid: Arco Libros, 1998.
- MARÍN N. *La enseñanza de las ciencias en primaria*. Almería: Grupo Editorial Universitario, 2003.
- MARTÍNEZ MARÍN. *Los diccionarios escolares del español: Lo que son y lo que deberían ser*. AHUMADA, I. (ed.). In: *Diccionarios españoles: contenidos y aplicaciones*. Jaén: Universidad de Jaén, p. 53-70, 1991.
- MARZANO, R.; PICKERING, D. *Building academic vocabulary. Teacher's manual*. Virginia: Association for Supervision and Curriculum Development, 2005.
- MALLART, J. *Competencias básica y enseñanzas mínimas del currículum de Lengua*. Revista de Educación 329, p. 219-298, 2002.

- MOON, R. *Sinclair, lexicography, and the Cobuild Project the application of theory*. International Journal of Corpus Linguistics 12, p. 159-181, 2007.
- ONDA. *Diccionari junior nou*. Barcelona: Onda, 2004.
- OXFORD UNIVERSITY PRESS. *Oxford First Illustrated Science Dictionary*, 2001.
- OXFORD UNIVERSITY PRESS. *Oxford First Illustrated Maths Dictionary*, 2001.
- OXFORD UNIVERSITY PRESS. *Oxford First Science Dictionary*, 2001.
- OXFORD UNIVERSITY PRESS. *Oxford First Maths Dictionary*, 2001.
- OXFORD UNIVERSITY PRESS. *Oxford Primary Illustrated Science Dictionary*, 2003.
- OXFORD UNIVERSITY PRESS. *Oxford Primary Illustrated Maths Dictionary*, 2003.
- OXFORD UNIVERSITY PRESS. *Student's Science Dictionary*, 2006.
- PÉREZ LAGOS, F. *Los diccionarios escolares de los últimos años: ¿una nueva lexicografía didáctica?*. ALVAR, M.; CORPAS, G. (cords.). In: *Diccionarios, frases, palabras*. Málaga: Universidad de Málaga, 1998.
- PIERCE, R. *Illustrated Mathematics Dictionary*, 2010. Disponible en: <<http://www.mathsisfun.com/aboutmathsisfun.html>>. Accedido en: setiembre de 2018
- PRADO, J. *El diccionario y la enseñanza de la lengua*. RUSCHTALLER, S.; PRADO, J. (eds.) In: *Tendencias en la investigación lexicográfica del español: el diccionario como objeto lingüístico y didáctico*. Huelva: Universidad de Huelva, p. 179-191, 2000.
- REY-DEBOVE, J. *Étude linguistique et sémiotique des dictionnaires français contemporains*. París-La Haya: Mouton, 1971.
- SANTILLANA. *Nou Diccionari escolar de la llengua catalana*. Barcelona: Santillana Grup Promotor, 2012.
- STOCKER, K. *Principios de didáctica moderna*. Buenos Aires: Kapelusz, 1964.
- TARP, S. *Funktionen in Fachwörterbüchern. Fachlexikographie. Fachwissen und seine Repräsentation*. BERGENHOLTZ H. et al. In: *Wörterbüchern*. Tübingen: Gunter Narr Verlag, p. 229-246, 1994.
- VOX. *Diccionari escolar de la llengua catalana*. Barcelona: Larousse Editorial, 2009.
- VYGOTSKY, L. *Pensamiento y lenguaje*. New York / Cambridge: Wiley and M.T.T. Press, 1962.
- WATT, M. et al. "Talking and writing for learning". BENTLEY, D.; WATTS, M. In: *Learning and teaching in school science*. Philadelphia: Open University Press, 1989.
- WESTPHAL, L. *Science Dictionary for kids*. Grades 4-9. Texas: Prufrock Press, 2009.

WILD, K. et al. *Oxford Children's Corpus*: A corpus of writing for children. Poster at ICLIC (International Corpus Linguistics Conference), Birmingham, UK, 2011.

WILD, K. et al. *The Oxford Children's Corpus*: Using a Children's Corpus in Lexicography". *International Journal of Lexicography* 26, 1, p. 190-218, 2013.

Como citar este texto (ABNT):

ESTOPÀ, R. Tradução de Tiele Kwarlevski. Construir para desconstruir e voltar a construir: elaboração colaborativa de um dicionário escolar de ciências. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 87-112, 2018.

INFORMAÇÕES CULTURAIS EM UM DICIONÁRIO BILÍNGUE DE APRENDIZAGEM¹

Michèle Fourment-Berni Canani²

Tradução: Samanta Siqueira³

Supervisão e revisão: Patrícia Reuillard⁴

Revisão técnica: Sandra Loguercio⁵

Resumo: O artigo apresenta as estratégias utilizadas em relação às informações culturais em uma obra de características particulares, já que se trata de um dicionário bilíngue de aprendizagem do francês para jovens itálofonos. Mostramos aqui como as informações, que não se limitam a apontar as diferenças culturais, mas apoiam-se na noção de cultura compartilhada, estão presentes em todos os níveis da microestrutura. Todos os exemplos, criados pela autora e constituídos de frases completas, são o lugar privilegiado para apresentar um quadro de vida cotidiana na qual atuam personagens, cuja fisionomia e relações se estabelecem na medida em que avançamos na consulta do dicionário. Um sistema de notas e de remissões, que se correlacionam nas diferentes partes da obra, esclarece e estende as informações fornecidas nos verbetes, criando um percurso de consulta no qual língua e cultura estão intimamente imbricadas.

Nos anos 1980-1990, Galisson notava “a renovação do interesse pela promoção das culturas entre os professores e pesquisadores cuja atividade – havia algumas décadas – era sobretudo promover as línguas estrangeiras” (1991, p. 111). Para ele, essa renovação era fruto de duas conjunturas: uma conjuntura interna, ligada à afirmação da abordagem comunicativa na didática das línguas, que transita necessariamente pelo “domínio das regras psicológicas, sociológicas e culturais – que condicionam o emprego da fala em situação” (*ibid.*); e uma conjuntura externa, ligada a condições históricas, políticas e

¹ Artigo publicado no periódico *Revue de didactologie des langues-cultures*, em 2002, disponível gratuitamente *online* na plataforma CAIRN. A presente tradução foi autorizada para ser publicada em português pela autora Michèle Fourment-Berni Canani. Referência bibliográfica completa do artigo original: FOURMENT-BERNI CANANI, M., Les informations culturelles dans un dictionnaire bilingue d'apprentissage, *Revue de didactologie des langues-cultures* 2002/4, N°128, p. 467-479. Disponível em: http://www.cairn.info/article.php?ID_REVUE=ELA&ID_NUMPUBLIE=ELA_128&ID_ARTICLE=ELA_128_0467. Último acesso em: 31 de outubro de 2018.

² Universidade de Roma “La Sapienza”.

³ Bacharela em Letras – Tradução Português/Francês, UFRGS.

⁴ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁵ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

ideológicas, que vão da evolução das ex-colônias em relação à língua e à cultura do opressor à forte imigração para o território francês que gera um “choque de culturas”.

O próprio termo “cultura”, carregado até então de um valor eminentemente literário, erudito, passa por um deslizamento semântico para designar tudo aquilo que “governa a maioria de suas [dos nativos] atitudes, de seus comportamentos, de suas representações, e dos costumes aos quais eles obedecem” (*ibid.*). A partir de então, a língua passa a ser tão estritamente associada – de maneira positiva – à cultura que um político, questionado sobre o problema da diversidade das línguas na Comunidade Europeia, respondia que “a variedade linguística da Europa é uma riqueza *cultural* evidente para todos os cidadãos europeus”⁶ e que pesquisadores lançam um sinal de alerta sobre o desaparecimento, em um ritmo vertiginoso, de certas línguas faladas no planeta, o que gera o apagamento dos conhecimentos que elas veiculam e a perda definitiva de partes inteiras de culturas (ver NETTLE, D.; ROMAINE, S., 2001).

É natural então que uma obra como o *Dizionario di apprendimento della lingua francese* (doravante DAF⁷, em formato impresso, destinado a jovens italófonos e concebido explicitamente para a aprendizagem de uma língua estrangeira, no caso, do francês, dê um espaço para essa dimensão. No entanto, o fato de que ele faça parte de uma categoria de obras normalmente previstas para a consulta e nunca associadas a uma função de aprendizagem – já que se trata de um dicionário bilíngue – torna necessária uma série de operações, levando em conta ao mesmo tempo o caráter específico do objeto dicionário, com suas convenções, e o texto didático, entendido como objeto de mediação⁸.

Se consideramos, como em nosso caso, que o binômio língua/cultura não implica uma justaposição – de um lado a língua, do outro a cultura – mas sim uma imbricação, uma interpenetração em todos os níveis de uma na outra, fica extremamente difícil estabelecer uma linha divisória entre as duas e isolar o que poderia ficar restrito à rubrica “informações culturais” no dicionário. Como mostramos em estudos anteriores (FOURMENT-BERNI CANANI, 1984, 1989), as estruturas linguísticas em si são reveladoras de uma certa visão de mundo, de uma certa concepção das relações interpessoais em uma dada comunidade e da dificuldade de dar conta disso em um outro código linguístico.

Assim, por exemplo, em italiano, um jogo sutil dos pronomes pessoais, em particular dos dativos, permite que se atribua um papel de primeiro plano à pessoa representada pelo pronome, mesmo quando ela não constitui o sujeito gramatical da frase. No enunciado *Gli è morto il padre*, o sujeito verdadeiro não é o sintagma nominal [*il*

⁶ Michel Rocard entrevistado por Fabio Gambino para o jornal *Repubblica*, 13 de julho de 2002. Grifo da autora.

⁷ Michèle Fourment (Berni Canani), *DAF, Dizionario di apprendimento della lingua francese*, Torino, Paravia, 1998.

⁸ Sobre a noção de mediação associada ao dicionário, ver FOURMENT-BERNI CANANI, “Le dictionnaire bilingue : de la médiation linguistique à la médiation didactique”, in: *La médiation et la didactique des langues et des cultures, Le Français dans le monde*, janeiro de 2003.

padre], mas o pronome dativo [*Gli*], colocado no início da frase; em francês, o melhor equivalente não é, em nossa opinião, *Son père est mort* [Seu pai morreu], mas *Il a perdu son père* [Ele perdeu seu pai]. Da mesma maneira, um emprego maciço do artigo definido em italiano onde o francês só pode empregar o indefinido ou o possessivo revela indicadores fixos, um quadro preestabelecido na esfera familiar e no meio próximo que permite um bom funcionamento da referência: enquanto o italiano pode sair com “a” noiva (*E’uscito con la fidanzata*) e comprar “o” carro (*Ha comprato la macchina*), o francês sai com “sua” noiva e compra “um” carro. Observamos também que o emprego frequente que o italiano faz em relação ao francês dos sufixos no nível do discurso, em particular de diminutivos, mostra uma tendência geral a favorecer o lado afetivo nas relações humanas, a minimizar os conflitos nas situações que podem ser duras, a atenuar os julgamentos que podem machucar, enfim, a despertar no outro uma recepção complacente: é só notar a frequência dos adjetivos depreciativos alterados em italiano, tais como “bruttina”, “antipaticuccia”, “volgarotto” etc.

Desse modo, quando desejamos, em um dicionário, mostrar ou mesmo colocar em evidência essas diferenças de estruturas discursivas entre as duas línguas, só podemos fazer isso grafando-as por meios tipográficos em exemplos criados como frases completas, como é feito no *DAF*⁹. É impossível, porém, a menos que se recorra a glosas específicas que tornariam esse tipo de obra muito pesado, revelar seu alcance cultural. É então em outros níveis que o linguista-lexicógrafo pode intervir em matéria de informações culturais.

1. Nível da macroestrutura

No *DAF*, a escolha das entradas não apresenta grandes variações em relação a um dicionário de francês impresso de dimensões similares destinado a usuários de língua materna da mesma faixa etária, ou seja, do nível do *collège* [correspondente ao 6º-9º ano do ensino fundamental no Brasil]. Consideramos que, para os dois tipos de leitores, o desenvolvimento tanto no nível linguístico quanto cognitivo é comparável. Sem dúvida, é um objetivo mais didático do que cultural que está na origem da decisão de colocar um artigo antes de cada substantivo. Isso permite de fato atender às duas exigências fundamentais que guiaram a elaboração do dicionário: aquela da contextualização (um substantivo é muito raramente empregado sem determinante) e aquela do contraste (já que as variações de gênero entre as duas línguas, mesmo próximas, são diversas vezes constatadas como fontes de erros).

Por outro lado, foi buscando facilitar e, em muitos casos, simplesmente permitir a observação de certas formas verbais correntes, que diferem de seu infinitivo (única forma apresentada na nomenclatura nas obras tradicionais), que foram introduzidas formas conjugadas de verbos de uso frequente como entradas, tais como: *je vais, nous avons, que nous ayons, nous prenons, je buvais, je verrai* [eu vou, nós temos, que nós tenhamos, nós

⁹ Os elementos que apresentam as diferenças de estruturas entre o italiano e o francês são marcados em negrito.

tomamos, eu bebia, eu verei] (esta última especialmente importante em razão do seu homógrafo italiano *verrai*, segunda pessoa do singular do futuro do verbo *venire* [vir]).

2. Nível da microestrutura

A concepção e a elaboração da microestrutura, por sua vez, não perderam de vista, em nenhum momento, o substrato no qual se inscreve a língua, substrato que, para um estrangeiro, pode se mostrar em uma cultura compartilhada ou ser completamente desconhecido. Não discutiremos a organização dos verbetes aqui, que não reproduz – como é normalmente o caso na produção lexicográfica bilíngue – aquela de um dicionário monolíngue, mas leva em conta, para a distribuição das diferentes acepções, a língua do usuário.

2.1 Cultura compartilhada

Dada a proximidade tanto geográfica quanto cultural da França e da Itália, exploramos uma competência cultural supostamente compartilhada que, diminuindo ou eliminando as distâncias quando possível, tende a tranquilizar o leitor em fase de aprendizagem. Assim, os exemplos do *DAF*, todos criados, são repletos de referências que constituem geralmente a bagagem cognitiva de um adolescente tanto italiano quanto francês. Certas entradas são a oportunidade para relembrar leituras que vão do *Chapeuzinho Vermelho*, na entrada *loup* [lobo], da *Cinderela*, na entrada *citrouille* [abóbora], dos Caninos Brancos, na entrada *adaptation* [adaptação], até *Robinson Crusoé*, na entrada do adjetivo *désert* [deserto], *Robin Hood*, na entrada *arc* [arco], Julio Verne, nas entradas *ballon* [balão]¹⁰ e *oeuvre* [obra]. Também as histórias em quadrinhos são apresentadas com o Tio Patinhas, nas entradas *obsession* [obsessão], *obséder* [obcecar] e *dollar* [dólar], *Asterix* na entrada *s'acharner* [obstinar-se], *Obelix* na entrada *obélisque* [obelisco], entre outros personagens. Outros autores e obras que fizeram ou provavelmente farão parte das aulas de francês são citados, como *O pequeno príncipe*, na entrada *prince* [príncipe], Molière na entrada *théâtre* [teatro], Proust na entrada *longtemps* [há muito tempo atrás] etc.

Essa vontade de evidenciar o que constitui um acervo cultural comum é reforçada pelo recurso das notas com perguntas diretamente dirigidas ao usuário, ou ainda por uma referência explícita à sua própria cultura. Assim, logo depois da entrada *passionner* [apaixonar], onde se encontra o exemplo “Os livros de Julio Verne apaixonam Jacques”, perguntamos: “Você já leu os livros do Julio Verne? Quais?”¹¹; na entrada *anneau* [argola], o exemplo proposto é “O campeão do mundo nas argolas é um italiano”¹².

¹⁰ Cf. fragmento textual n.1 no fim deste artigo.

¹¹ Como tudo que constitui a metalinguagem está em italiano (língua do destinatário do dicionário), nós fornecemos uma tradução.

¹² N.T.: Referência ao ginasta italiano Jury Chechi, que participou de várias olimpíadas na modalidade de ginástica artística masculina.

Outros aspectos relativos a eventos conhecidos ou tradições comuns são também apresentados, como o feito dos astronautas na entrada *lune* [lua], a comemoração de finados no dia 2 de novembro na entrada *cimetière* [cemitério], a árvore de Natal e o presépio na entrada *Noël* [Natal] etc. Ocupações ou lazeres que todos os adolescentes conhecem, como futebol, pesca, cinema e teatro são normalmente evocados.

2.2 Diferenças culturais

Falamos de diferença cultural em todos os casos em que a palavra-entrada não encontra equivalente em uma língua-cultura de chegada ou quando a tradução remete, no nível pragmático, a uma realidade diferente.

As lexias que não têm equivalentes em outra língua são normalmente relativas a campos específicos como, por exemplo, o da gastronomia. O recurso ao empréstimo é então adotado com o complemento de uma glosa entre parênteses. É o caso, entre outros, da palavra *camembert*: a glosa (“queijo típico francês”) remete ao protótipo “queijo” que, este sim, é conhecido.

Não falaremos aqui das palavras – bastante numerosas – cujas polissemias não se sobrepõem e cujas produções fraseológicas seguiram percursos diferentes; nós nos limitaremos a abordar aquelas que, tendo um equivalente no plano lexical, remetem todavia a uma realidade pouco conhecida.

O caso da entrada *marée* [maré] é significativo¹³. No plano denotativo, o equivalente italiano *marea* é completamente satisfatório; no entanto, o conhecimento desse fenômeno no plano pragmático varia de maneira considerável se vivemos mais perto do Oceano Atlântico ou do Mar Mediterrâneo. Assim, o exemplo foi criado a fim de mostrar ao usuário que desconhece as atividades lúdicas ligadas aos movimentos ascendentes e descendentes do mar (“Quando a maré está baixa, Jeanne se diverte procurando conchas, quando a maré está alta, ela toma banho de mar”). Além disso, uma glosa, introduzida por um sinal tipográfico específico na sequência do verbete¹⁴, traz uma complementação à informação de maneira explícita: “nas costas francesas do Oceano Atlântico e do canal da Mancha, o fenômeno das marés é muito mais sensível do que no Mediterrâneo e regula o ritmo da vida”¹⁵.

Um terceiro caso pode ser ilustrado pela entrada *département* [departamento]: se a tradução *dipartimento* pode ser correta, ela é, no entanto, insuficiente para dar conta de

¹³ Cf. fragmento textual n.2 no fim deste artigo.

¹⁴ O DAF se caracteriza por um procedimento de notas distribuídas em diferentes rubricas, cada uma sendo introduzida por um sinal tipográfico diferente: um triângulo vermelho assinala uma dificuldade de ordem fonética, um perigo de interferência lexical ou sintática, a diferença de extensão de uma acepção; uma mãozinha vermelha com o indicador apontado remete a outras entradas do dicionário para permitir, entre outras coisas, um uso não somente passivo (limitado à compreensão), mas também ativo (estendido à produção) da língua francesa; um “i” em um quadrado branco introduz uma informação cultural.

¹⁵ Em contrapartida, um outro fenômeno natural, por exemplo o arco-íris, que pertence a uma cultura compartilhada, não passa por um tratamento particular e apresenta simplesmente um exemplo contrastivo.

uma divisão administrativa própria à França. Aqui, o exemplo tem uma função informativa (“A França conta com 96 departamentos mais 5 [sic] ultramarinos) e, como antes, é reforçado por uma nota ao final do verbete: “A França, no nível administrativo, é dividida em *départements*”.

Isso nos leva ao problema dos falsos cognatos, particularmente numerosos quando somos confrontados com duas línguas-culturas tão próximas como o italiano e o francês e que só podem ser tratados de maneira séria, se levarmos em conta a dimensão diacrônica de cada uma dessas línguas, as restrições, os aumentos ou deslocamentos de sentido.

Os empréstimos representam um caso particular. Muitos termos de moda, cozinha, mobiliário do léxico italiano são decalques mais ou menos adaptados ao código fonético ou gráfico da língua receptora; mas muito raros são aqueles que não sofreram deslizamento semântico. Assim, a palavra *abat-jour* designa em italiano atual uma luminária com uma luz difusa, e não a parte da luminária que cobre a fonte luminosa, a pantalha. O exemplo que considera essa divergência, fonte de erros (“A luz é ofuscante porque não tem pantalha sob a lâmpada”), é reforçado por uma nota explícita ao final do verbete (“O italiano, que tomou emprestado essa palavra do francês, emprega-a no sentido de “luminária com pantalha” que, em francês se diz *lampe* [luminária]”) e por um desenho que delimita de maneira clara o objeto e indica sua função. Um tratamento similar é reservado a entradas como *boutique* [loja].

2.3. Informações culturais

Essas informações culturais não se referem diretamente à palavra-entrada, mas são distribuídas através das frases-exemplos por pinceladas sucessivas que aos poucos, na medida em que se avança na consulta do dicionário, formam uma tela de fundo ou, mais exatamente, uma paisagem com características bem particulares. Elas podem abarcar aspectos geográficos ou históricos: “O rio Marne é um afluente do Sena”, na entrada *affluent* [afluente]; “Toulouse fica a cerca de 250 quilômetros de Bordeaux”, na entrada *à peu près* [cerca de] que, no final do verbete correspondente, sugere a consulta de um mapa para situar as duas cidades; “Você sabe em que ano aconteceu a Revolução Francesa?”, na entrada *révolution* [revolução]; “Durante a Revolução Francesa, o carrasco guilhotinava os condenados à morte”, na entrada *bourreau* [carrasco], com uma nota explicativa sobre o instrumento e a origem de seu nome. O dia 14 de julho é várias vezes citado em verbetes de palavras como *bal* [baile] ou *défilé* [desfile] e seu *status* de data nacional relacionado à Tomada da Bastilha é comentado em notas.

A alusão ao galo como símbolo da França aparece na entrada homônima; a palavra *devise* [lema] é ilustrada por “O lema da França é: Liberdade, Igualdade, Fraternidade”. Os jogos (batalha, cabra-cega e quatro cantos), as manifestações esportivas e culturais mais conhecidas (o Tour de France, o Festival de Cannes – que permitem também inserir uma noção geográfica) também são citados.

Uma grande parcela das informações é reservada às tradições: o lírio-do-vale no Dia do trabalhador na entrada *brin* [ramo], os crepes na Terça-feira Gorda, o bolo-rei no

Dia de Reis, acompanhado de uma nota contrastiva sobre a personagem folclórica italiana de mesmo dia, Befana, entre outras.

Considerando o público-alvo, uma atenção particular foi dada às informações relativas ao sistema escolar, experiência compartilhada, mas de maneiras e ritmos diferentes. Para as entradas que se referem diretamente à organização e à distribuição do ensino, ora o exemplo com sua tradução em italiano é suficiente, como no caso de *la sixième* [sexto ano do fundamental], que corresponde à *prima media* em italiano, de *la cinquième* [sétimo ano do fundamental], que corresponde à *seconda media* e assim por diante, ora uma breve nota no final do verbete assinala uma diferença. É o caso do termo *lycée*, que, contrariamente ao *liceo* italiano, abarca tanto o ensino médio clássico quanto o técnico e profissional, ou ainda o *collège* [correspondente aos últimos anos do ensino fundamental brasileiro], que, tendo o seu falso-cognato italiano *collegio*, exige um esclarecimento.

O sistema de notas, diferente nos dois países, é explicitado, e a acepção particular da palavra *moyenne* [média] no contexto escolar e sua correspondente italiana *sufficienza* é claramente distinguida no exemplo e na nota da outra acepção mais geral que equivale à *media*. Sinalizamos que as férias que são chamadas de “grandes” em francês, são normalmente de “verão” na Itália; que as férias de início de primavera não coincidem necessariamente com as férias de Páscoa (exemplo fornecido na entrada do verbo *coïncider* [coincidir] e complementado por uma nota); que nas quartas-feiras os alunos geralmente não vão à escola na França e se dedicam a atividades extraescolares (exemplo na entrada *mercredi* [quarta-feira]: “Na quarta-feira de tarde, Marie toca flauta” e especificação em uma nota).

Além disso, pareceu-nos importante lembrar certos aspectos específicos em relação às matérias ensinadas e que contribuem diretamente para a formação cultural. Assim, por exemplo, o exercício de dissertação que, através do exemplo fornecido na entrada homônima, distingue-se da redação, da qual se aproxima o exercício da escrita equivalente em italiano, *tema*, para melhor distingui-lo do exercício de *thème* [versão, traduzir da língua materna para a língua estrangeira] que diz respeito a uma atividade totalmente diversa. O lugar preponderante ocupado pela ortografia e suas dificuldades é muitas vezes sugerido com o intuito de dar conta de um traço que caracteriza a língua francesa e de tranquilizar o aprendiz estrangeiro que se confronta com isso, mostrando os obstáculos que os alunos franceses também encontram.

2.4 Informações metalinguísticas

Buscamos também, sempre que possível para não sobrecarregar a obra, abrir espaço para certos elementos que poderiam contribuir para a consciência do que constitui o corpo da língua, para a maneira como os significados se estratificam, para as relações entre o sentido literal e o sentido figurado. Em suma, buscamos suscitar uma reflexão que nos parece essencial sobre a outra língua e, portanto, sobre a própria língua, e lançar as bases de uma cultura linguística. Uma breve nota assinala que a haste de um guarda-chuva

ou de um espartilho, chamada de *baleine* [baleia] tem seu nome do mamífero referido na entrada, extensão que o italiano não fez. O fato de que a flor *coquelicot* [papoula] deve seu nome à sua similaridade com a crista do galo [*coq*, em francês] – como é ressaltado – é certamente instrutivo e facilita a memorização.

Sinalizamos, nas respectivas entradas, os nomes próprios que originavam nomes comuns, mostrando sua motivação (o arquiteto Mansart para *mansarde* [mansarda]; o prefeito Poubelle para a *poubelle*, [lata de lixo]); mas também o fenômeno inverso como o do bairro parisiense Marais [pântano], que tem seu nome devido à zona pantanosa em que fica situado¹⁶.

Também tratamos com um cuidado especial as palavras da metalinguagem, sobretudo quando apresentam dificuldades de interpretação se comparadas com o italiano. Assim, a entrada *alphabet* [alfabeto] possibilita, através do exemplo “O alfabeto francês tem mais letras do que o alfabeto italiano” e da sugestão em nota para a consulta ao alfabeto francês no início do livro, sinalizar que cada língua tem suas próprias convenções. O *attribut* [atributo], em sua acepção gramatical, suscita muita estranheza por parte de um italo-fono que o designa pela palavra *predicato*. A função do exemplo “Na frase ‘Marie é loira’, ‘loira’ é o atributo do sujeito ‘Marie’” é eliminar a ambiguidade. Os diferentes acentos gráficos do francês, que constituem uma dificuldade para todo aprendiz estrangeiro, recebem entradas autônomas que são retomadas por um desenho recapitulativo.

2.5 Personagens

Os exemplos¹⁷, que constituem a estrutura da obra, trazem personagens que transitam no mundo da escola e no ambiente familiar: cada um com seu nome, suas características físicas, psicológicas e sociais, que se estabelecem na medida em que avançamos na consulta do dicionário e com as quais os alunos adolescentes podem se identificar.

A escolha dos nomes constitui uma primeira informação sobre a diversidade étnica da França hoje: Marie, Philippe, Caroline, Yasmine, Jeanne, Chakib, Jim. Se cada um apresenta características próprias de um país, de uma categoria social, de um pertencimento religioso, procuramos dar a eles uma personalidade mais complexa, que ultrapasse o estereótipo e os faça indivíduos¹⁸. *Marie* é uma criança loira que vive em um meio favorecido, mimada por seus pais; *Chakib* não é somente um menino argelino de família muçulmana (“A mãe de Chakib faz muito bem o cuscuz” na entrada *couscous* [cuscuz], “Como muitas mulheres muçulmanas, a mãe de Chakib usa véu”, na entrada

¹⁶ Cf. fragmento textual n.3 no fim deste artigo.

¹⁷ Assinalemos que os cerca de 24.000 exemplos do *DAF*, todos criados, empregam somente palavras que figuram na nomenclatura do dicionário.

¹⁸ Lembremos que a introdução de personagens em um dicionário para crianças já tinha sido feita por Josette Rey-Debove no *Le petit Robert des enfants*, escolha claramente original e inovadora na época da publicação da obra, em 1988.

voile [véu]), apaixonado por *Marie* que o rejeita (“Às vezes Marie diz palavras ofensivas a Chakib”, na entrada *blessant* [ofensivo]): além de suas características físicas (“Chakib tem cabelos castanhos e crespos”, na entrada *brun* [castanho]), sabemos que ele adora a natureza e os animais (“Chakib gostaria muito de viver no interior”, na entrada *aimer* [gostar] para ilustrar o sintagma “gostar muito”; “Chakib gosta dos bichos; ele tem um gato, um cachorro, um papagaio e um *hamster*”, na entrada *bête* [bicho]), que ele queria ser veterinário, que é levado (“Chakib adora contar mentiras no 1º de abril”, na entrada *avril* [abril]), que é briguento, que tem uma irmãzinha de quem cuida bastante etc. *Yasmine* tem uma saúde fraca, tem sérios problemas de ortografia (“Yasmine é ruim em ortografia; ela errou de novo dez vezes no ditado”, na entrada *dictée* [ditado]) e para acompanhar o planejamento escolar em geral (“Os pais de Yasmine esperam que ela não rode”, na entrada *redoubler* [rodar]).

Todos esses traços, dos quais só podemos citar uma ínfima parte, encontram-se e coincidem em diferentes momentos e atividades ilustradas nos exemplos para formar um mural, uma passagem de vida.

2.6 Ilustrações

As ilustrações constituem um traço tipológico dos dicionários destinados aos adolescentes. Longe de constituir simples enfeites, elas desempenham funções múltiplas e vêm reforçar informações contidas no texto. Elas se apresentam em nosso dicionário de três formas: quadros temáticos, desenhos no início dos capítulos para cada letra do alfabeto e ilustrações pontuais no corpo da obra.

Além dos quadros geográficos no final do volume (a França e a Itália com suas regiões, os cinco continentes com indicação dos principais países), o dicionário apresenta quadros sintagmáticos, de composição clássica, que agrupam em torno de um tema – a cidade, o mar, o interior etc. – objetos apresentados em seu meio, dentre os quais tivemos o cuidado de introduzir aqueles que trazem hábitos específicos do país estrangeiro (assim, na ilustração destinada ao quarto, mostramos um *traversin*, [travesseiro longo para a cabeceira], que não é ou é pouco usado na Itália e cuja tradução por meio de uma perífrase correria o risco de continuar opaca).

Os desenhos que acompanham a letra inicial de cada capítulo representam alguns objetos cujo nome em francês começa pela letra em questão enquanto seu equivalente italiano está longe disso (para a letra *b*: *berceau* [berço], *bouchon* [rolha], *baguette* [baguete], *balai* [vassoura]; para a letra *j*: *journal* [jornal], *jongleur* [malabarista], *jouet* [brinquedo], *jupe* [saia] etc.).

Quanto às ilustrações presentes ao longo da obra, elas foram criadas com o objetivo de sustentar e de prolongar a função metalinguística dos exemplos, sendo parte integrante do texto. Estão vinculadas a entradas nominais e chamam a atenção para problemas de polissemia que não se sobrepõem entre as duas línguas, já claramente evidenciados tanto pela estruturação do verbete quanto pela escolha dos exemplos: ora trata-se de uma palavra francesa polissêmica, como *ampoule* [lâmpada] (em italiano: *lampadina*, *fiala*,

vescica) ou *aiguille* [agulha] (em italiano: *ago*, *lancetta*), cujas diferentes acepções são representadas pelos objetos correspondentes agrupados em um só desenho, ora trata-se de palavras francesas diferentes para nomear objetos que o italiano designa por um só vocábulo, como é o caso de *échelle-escalier* [escada extensiva].

Elas chamam atenção para simetrias enganosas, já que parciais, pontuando as diferenças de emprego na língua estrangeira entre palavras como *gare* [estação de trem] e *station* [estação] (para a mesma palavra italiana *stazione*). A proximidade formal de certas unidades que podem engendrar assimilações enganosas justifica os desenhos que evidenciam a diferença entre um *immeuble* [imóvel] e um *palais* [palácio] (respectivamente em italiano *palazzo* e *palazzo sontuoso* ou *reggia*), um *cadre* [quadro, no sentido de “panorama”, “representação”, “moldura”] e um *tableau* [quadro artístico ou escolar] (respectivamente em italiano *cornice* e *quadro*), ou ainda a ilustração agrupando uma *cantine* [cantina] e uma *cave* [adega ou porão], chamando a atenção, assim, para o falso cognato *cantina* que equivale ao francês *cave*. Além disso, um objeto de uso cotidiano durante o café da manhã de um jovem francês como o *bol* [tigela], pouco conhecido na cultura receptora e cujo equivalente *scodella* fica ambíguo, já que pode designar também um prato fundo, é representado ao lado de uma xícara, objeto, pelo contrário, familiar.

Conclusão

Se, como propõe Robert Galisson, organizamos sob o termo de “cultural” o que tange à cultura cotidiana – a saber, um mínimo de conhecimentos compartilhados por uma comunidade linguística que permite a seus membros criar entre si certas relações de convivência –, é natural que um dicionário de aprendizagem do francês destinado a estrangeiros só possa fundir os dois elementos – o linguístico e o cultural – em uma alquimia cujos ingredientes são difíceis de separar.

Pensamos que a cultura, como a língua, é um dos componentes essenciais da competência comunicativa, mas que essa competência não poderia excluir a reflexão metalinguística, ou seja, a capacidade de refletir sobre a língua durante a aprendizagem e, em contrapartida, sobre a sua própria língua. Nesse sentido, a abordagem contrastiva, tanto na elaboração dos verbetes quanto na criação dos exemplos e das ilustrações, pondo em questão um recorte do real, uma visão do mundo e um sistema de valores que poderiam se revelar únicos e evidentes, constitui o melhor acesso à noção de relativismo cultural.

Exemplo 1

un ballon [uma bola] nome 1 *un pallone*. *Le ballon de foot est rond tandis que le ballon de rugby est ovale* [A bola de futebol é redonda, enquanto que a de rugby é oval], *il pallone da calcio è rotondo mentre il pallone da rugby è ovale*. 2 *un pallone aerostatico*. *Jules Verne a écrit “Cinq semaines en ballon”*, [Julio Verne escreveu “Cinco semanas em um balão”], *Jules Verne ha scritto “Cinque settimane in pallone”*. 3 *un palloncino*. *Mon petit frère a*

laissé s'envoler son ballon [Meu irmãozinho deixou voar seu balão], il mio fratellino ha lasciato volar via il suo palloncino.

[i] Jules Verne è il famoso autore di *Vingt Mille Lieues sous les mers* (“Ventimila leghe sotto i mari”) e di *Le Tour du monde en quatre-vingts jours* (“Il giro del mondo in ottanta giorni”); hai letto questi o altri libri di Jules Verne?

Exemplo 2

la marée nome : la marea. *À marée basse Jeanne s’amuse à chercher des coquillages, à marée haute elle se baigne* [Na maré baixa, Jeanne se diverte catando conchinhas, na maré alta, ela entra no mar], quando la marea è bassa Jeanne si diverte a cercare delle conchiglie, quando la marea è alta fa il bagno.

[i] sulle coste francesi dell’oceano e della Manica il fenómeno delle maree è molto più vistoso che nel Mediterraneo e regola il ritmo della vita.

Exemplo 3

un marais nome 1 una palude. *Il y a beaucoup de moustiques parce que nous sommes près d’un marais* [Há muitos mosquitos porque estamos perto de um pântano], ci sono molte zanzare perché siamo vicini a una palude. 2 *un marais salant* [uma salina], una salina.

[s] *un marécage*

[i] le Marais è il nome di un quartiere nel centro di Parigi che un tempo era una palude.

Referências

CORMIER Monique, OUMET Catherine, BOULANGER Jean-Claude. 2001. « À propos de la néobienséance dans les dictionnaires scolaires : les prénoms dans les exemples » in *Les dictionnaires de langue française*. Paris : Honoré Champion.

FOURMENT-BERNI CANANI, Michèle. 1984. « De la détermination en français et en italien », *Revue de linguistique romane*, n° 189-190, Strasbourg.

_____. 1989. « Les clitiques datifs et leur fonction “dramatique”. Étude comparative italien-français », *SILTA*, XVIII-3.

_____. 2000. « La conception d’un dictionnaire d’apprentissage du français pour italophones » in *Approches contrastives en lexicographie bilingue*. Paris : Honoré Champion.

_____. « Structures linguistiques à portée culturelle dans les dictionnaires italien-français/français-italien » in *Le traitement des écarts culturels, Actes des IIIe journées d’étude sur la lexicographie bilingue (INALCO 12-13 octobre 2000)*. Paris : Honoré Champion.

_____. « Le dictionnaire bilingue : de la médiation linguistique à la médiation didactique » in *La médiation et la didactique des langues et des cultures*, n° spécial *Le Français dans le monde*, janvier 2003.

GALISSON, Robert. 1988. « Pour une politique de rénovation des dictionnaires monolingues », *Reflète*, n° 27.

_____. 1991. *De la langue à la culture par les mots*. Paris : CLE International, Collection Didactiques des langues étrangères.

_____. 1995. *Lexiculture et enseignement*, ÉLA, n° 97.

LEHMANN, Alise. 1991. « Une nouvelle conception du dictionnaire d'apprentissage : le Petit Robert des enfants », *Cahiers de lexicologie*, vol. 59, fasc. 2.

_____. 1993a. « L'exemple et la définition dans les dictionnaires pour enfants », *Repères*, n° 8 [nouvelle série]. [Pour une didactique des activités lexicales à l'école].

_____. 1993b. « Le renouvellement de la lexicographie contemporaine dans le dictionnaire d'apprentissage : le cas du Robert des jeunes » in *Actes du XXe Congrès international de linguistique et philologie romanes*, Université de Zurich, 6 au 11 avril 1992. Tübingen : A. Francke Verlag, tome 4.

_____. 2000. « Les dictionnaires pour enfants : diversité et uniformisation », *Construire les compétences lexicales, Le Français aujourd'hui*, n° 131.

LEVY-MONGELLI, Danielle. 1993. « Sur la notion de progression dans la construction scolaire d'une compétence culturelle en langue étrangère : quelques critères », *Heteroglossia*, n° 5. Ancona : Nuove Ricerche.

MARGARITO, Mariagrazia. 2001. « Italianismes du français. Notes sur des parcours de recherche », *Cahiers de lexicologie*, n° 78, p. 117-126.

NETTLE Daniel, ROMAINE Suzanne. 2001. *Voci del silenzio, sulle tracce delle lingue in via di estinzione* (trad. italiana). Roma : Carocci editore.

PRUVOST, Jean (sous la direction de). 2001. *Les dictionnaires de langue française*. Paris : Honoré Champion.

REY, Alain. 1986. « Les écarts culturels dans les dictionnaires bilingues », *Lexicographica*, 2. Tübingen : Max Niemeyer Verlag.

_____. 1987a. « La notion de dictionnaire culturel et ses applications », *Cahiers de lexicologie*, n° 51.

_____. 1987b. « Le dictionnaire culturel », *Lexicographica*, n° 3. Tübingen : MaxNiemeyer Verlag.

REY-DEBOVE, Josette. 1989. « Dictionnaires d'apprentissage : que dire aux enfants ? » in *Lexiques*, coordonné par Amr Helmy Ibrahim, coll. « Recherches et Applications ». Paris : Hachette.

_____. (rédaction dirigée par). 1990. [1988]. *Le petit Robert des enfants*. Paris : Dictionnaires Le Robert.

_____. 1993. « Le contournement du métalangage dans les dictionnaires pour enfants : translation, monstration, neutralisation », Repères, n° 8 [nouvelle série]. [Pour une didactique des activités lexicales à l'école].

Como citar este texto (ABNT):

FOURMENT-BERNI CANANI, M. Tradução de Samanta Siqueira. Informações culturais em um dicionário bilíngue de aprendizagem. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 113-125, 2018.

CARACTERÍSTICAS DISTRIBUTIVAS DO ESPANHOL DE CUBA A PARTIR DA ANÁLISE DE DOIS ESTUDOS LINGÜÍSTICOS¹

Lucía Marconi², Leonel Ruiz Miyares³, Paola Cutugno⁴

Tradução: Cyrano da Rosa Silva⁵

Supervisão e revisão: Cleci Regina Bevilacqua⁶

1. Introdução

Desde 1998, existe uma frutífera colaboração acadêmica entre o grupo de pesquisas do Instituto di Linguistica Computazionale “Antonio Zampolli” (ILC, Gênova), pertencente ao *Consiglio Nazionale delle Ricerche* (Itália), e o Centro de Linguística Aplicada (CLA), do Ministério de Ciência, Tecnologia e Meio Ambiente, em Santiago de Cuba. Essa colaboração, em particular com a Dra. Eloína Miyares Bermúdez, com o Dr. Julio Vitelio Ruiz Hernández, com o Dr. Leonel Ruiz Miyares e com outros pesquisadores do CLA, ocorreu por meio da construção conjunta de duas obras editoriais: o *Diccionario Ortográfico de Español* (DOE) (MARCONI, MIYARES BERMÚDEZ, RATTI, ROLANDO e RUIZ MIYARES, 2000a) e o *Vocabulário Inverso e Anagramas do Espanhol* (VIAE) (MARCONI, MIYARES BERMÚDEZ, RATTI, ROLANDO e RUIZ MIYARES, 2000b).

As entradas do DOE e do VIAE foram obtidas a partir dos dados da obra científica *Léxico Activo-Funcional del Escolar Cubano* (LAFEC) (MIYARES BERMÚDEZ, 2006) realizada por pesquisadores do CLA, de Santiago de Cuba, sob a direção de Eloína Miyares.

Conforme destacam Ruiz Hernández e Miyares Bermúdez (1997, p. 67):

Os pesquisadores do nosso Centro, desde a sua fundação, há exatamente vinte e cinco anos, se dedicaram a realizar pesquisas

¹ A presente tradução foi autorizada para ser publicada em português por Leonel Ruiz. Referência bibliográfica completa do artigo original: MARCONI, L. RUIZ MIYARES, L. CUTUGNO, P. Características distributivas del español de Cuba a partir del análisis de dos estudios lingüísticos. In: MIYARES, L. R. (ed.) *Estudios de Lexicología y Lexicografía. Homenaje a Eloína Miyares Bermúdez*. Santiago de Cuba, Ediciones Centro de Linguística Aplicada, 2018, p. 87-110.

² Instituto di Linguistica Computazionale “Antonio Zampolli”, Consiglio Nazionale della Ricerche Gênova, Italia.

³ Centro de Linguística Aplicada, Ministério de Ciência, Tecnologia e Meio Ambiente, Santiago de Cuba, Cuba.

⁴ Instituto di Linguistica Computazionale “Antonio Zampolli”, Consiglio Nazionale Dell e Ricerche Gênova, Italia.

⁵ Aluno do Curso de Bacharelado em Letras – Tradução Português/Espanhol, UFRGS.

⁶ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

aplicadas com a finalidade de proporcionar aos professores do ensino básico, aos professores universitários, aos metodólogos e a outros trabalhadores da educação básica, bem como aos meios de difusão, especialmente o rádio e a televisão, alguns resultados científicos que permitissem o conhecimento das características fundamentais do espanhol de Cuba, especialmente no plano fonético e no nível léxico da língua. Para isso, realizamos algumas pesquisas de caráter descritivo da pronúncia de nossos falantes e, posteriormente, do vocabulário ativo dos estudantes do país, sendo concluídas em 1975 e 1995, respectivamente. Em ambas as vertentes de pesquisa, depois de coletar os resultados referidos, tivemos a tarefa de realizar trabalhos de aplicação e generalização dos dois níveis mencionados anteriormente – plano fonético e o nível léxico – incluindo o importante aspecto ortográfico.

O LAFEC é um dos resultados das pesquisas do CLA que permitiu o conhecimento de algumas características fundamentais do espanhol de Cuba. Essa obra foi feita através do processamento de textos de estudantes de ambos os sexos, da 2ª a 6ª série, dos setores urbanos e rurais em todo o país. Na pesquisa, foram coletadas 5.800 composições escritas e 1.200 amostras gravadas, e, ao final, os dados coletados constituíram um *corpus* de mais de 700.000 palavras, das quais mais de 25.000 são diferentes. Essas composições foram escritas a partir de diversas temáticas que cobriam esferas como a natureza, família, escola, entretenimento, entre outras, que permitiram uma aproximação à realidade do vocabulário das crianças da escola primária cubana e, além disso, a identificação das principais dificuldades de ortografia apresentadas por elas.

Com as palavras diferentes do LAFEC e suas diversas categorias gramaticais, foi iniciado o trabalho para a elaboração do DOE, formado por mais de 8.400 entradas de vários tipos: verbos, substantivos, adjetivos, entre outras categorias.

Na parte inicial do livro, são reproduzidas as atuais regras ortográficas, as quais devem ser seguidas para se escrever corretamente. Em seguida, constam algumas informações sobre as irregularidades gráficas de alguns verbos e outras palavras. Logo após, há uma pequena descrição do dicionário, seguida pelos modelos de conjugação verbal e, finalmente, o dicionário com suas entradas. Cada entrada contém informações relativas à categoria gramatical, à flexão da palavra para substantivos e adjetivos e, para os verbos, indica-se o número que remete ao modelo de conjugação correspondente. Com os resultados do LAFEC, elaborou-se também, como já mencionado, o *Vocabulário Inverso y Anagramas del Español* (VIAE).

O DOE e o VIAE se complementam a partir dos elementos de flexão das categorias como substantivos, adjetivos, etc., incluídos no primeiro, e mais de 80.000 palavras flexionadas em ordem reversa e mais de 10.000 anagramas que constam no segundo. Essas

duas obras são, atualmente, instrumentos muito úteis de trabalho e consulta para estudantes e professores em todo o país.

No VII Simpósio Internacional Comunicação Social, realizado em Santiago de Cuba, em 2001, foi apresentado o estudo *Rimas y anagramas en el Léxico del Escolar, Cubano* (MARCONI, MIYARES BERMUDEZ, RATTI, ROLANDO, 2001), em colaboração com o CLA, sobre as terminações dos lemas e das palavras flexionadas e sobre os anagramas dos espanhol de Cuba.

Em 1995, para o italiano, foi realizado o trabalho *Radiografia dell'italiano* (RATTI, MARCONI, BURANI, 1995), baseado em dados de dicionário *Zingarelli minore* (ZINGARELLI, ZANICHELLI, 1987) e nas flexões geradas a partir das entradas do dicionário *Flessioni-Rime-Anagrammi* (RATTI, MARCONI, MORGAVI, ROLANDO, 1988). Nesse estudo, foram apresentados alguns dados sobre a distribuição dos lemas e das palavras flexionadas.

Na introdução da obra *Radiografia dell'italiano* especifica-se:

Frequentemente, no decorrer de nosso trabalho, nos perguntamos sobre os componentes formais da linguagem, seja do tipo qualitativo ou quantitativo, e percebemos que, em muitas áreas da ciência, onde a linguagem é objeto ou sujeito, há um crescente interesse nas propriedades de distribuição de unidades linguísticas pertencentes a um idioma específico.

As "propriedades de distribuição" expressam um conjunto de medidas empíricas que levam em conta o número de palavras de um tipo particular ou de uma categoria gramatical específica, pertencentes a um dado idioma e as proporções relativas, a frequência de ocorrência em comparação com todo o conjunto, o número absoluto e a frequência de certas unidades sublexicais, tais como sequências de letras, sílabas e morfemas.

A importância atribuída às propriedades distributivas nasce do interesse de pesquisar quais os componentes, nos processos de acesso a representação mental do léxico e nos processos linguísticos em geral, são atribuídos às propriedades universais da linguagem e ao sistema cognitivo e quais podem ser considerados dependentes de linguagem específica a qual um leitor está exposto⁷.

Analisar as propriedades distributivas significa, portanto, realizar uma série de medidas empíricas de uma dada língua, considerando, por exemplo, o número de palavras de um tipo específico ou de uma categoria gramatical de um determinado conjunto, seja um dicionário ou um léxico de frequências. Isso significa calcular as proporções, a

frequência de ocorrência, relacionando-a ao conjunto e ao grupo específico, bem como calcular a frequência de certas unidades sublexicais, como sequências de letras.

O objetivo deste artigo é realizar um estudo sobre as características distributivas de um subconjunto do espanhol de Cuba. Referimo-nos à informação contida no DOE – excluindo as lexias complexas e as locuções –, ou seja, aos 7.927 lemas gerados a partir das 93.759 palavras flexionadas com suas categorias gramaticais. Os dados do LAFEC também serão analisados para mapeamento das palavras utilizadas pelos alunos. A partir do LAFEC, foram extraídas as frequências de certas unidades sublexicais e, neste caso, foi feita referência apenas aos termos usados pelas crianças em suas produções e, posteriormente, às formas utilizadas.

O artigo também contém uma série de comparações de algumas propriedades distributivas derivadas do tratamento dos dados do DOE com os resultados obtidos da análise do LAFEC.

Quando se fala de propriedades distributivas de uma língua, essas podem ser consideradas em relação aos lemas, às várias palavras flexionadas ou ao número de ocorrências⁸ em um *corpus* específico. Neste artigo, faremos referência a uma amostra de lemas e flexões derivadas do DOE e somente às flexões do LAFEC.

2. A informação linguística e a estrutura quantitativa do léxico

Decidiu-se realizar uma série de cálculos sobre o espanhol de Cuba com os dados do DOE, para os quais foram consideradas todas as formas possíveis ou palavras flexionadas geradas a partir dos lemas. Na Tabela 1, encontra-se a distribuição dos lemas e das formas das diferentes categorias gramaticais, incluindo as palavras homógrafas.

<i>categoria gramatical</i>	<i>lema</i>	<i>forma</i>
adjetivo	1831	6486
advérbio	265	265
artigo	4	7
conjunção	14	14
contração	2	2
interjeição	27	27
preposição	17	17
pronome	59	132
sustantivo	4460	9471

⁸ Ocorrência do lema ou palavra flexionada (forma ou flexão) em um *corpus*, ou seja, quantas vezes é encontrada a palavra como lema ou na sua forma flexionada.

verbo	1248	77336
Total	7927	93759

Tabela 1 – Distribuição dos lemas e das formas das diferentes categorias gramaticais a partir de dados do DOE.

Nesse estudo de distribuição de lemas e de palavras flexionadas, as palavras homógrafas foram analisadas uma só vez; portanto, de 7.927 lemas foram analisados 7.535 lemas diferentes e de 93.759 formas, 80.336 diferentes.

A tabela 2 contempla a informação relacionada com a quantidade de lemas e palavras flexionadas presentes no DOE, considerando a quantidade de letras (longitude) da palavra.

<i>longitud e</i>	<i>lemas</i>	<i>formas</i>	<i>longitude</i>	<i>lemas</i>	<i>formas</i>	<i>longitude</i>	<i>lemas</i>	<i>formas</i>
1	5	5	9	958	13505	17	3	41
2	34	48	10	630	12858	18	3	8
3	95	184	11	371	9781	19	0	1
4	431	1085	12	196	6423	20	0	0
5	884	3041	13	94	3617	21	1	2
6	1150	5619	14	61	1623	22	0	2
7	1301	9344	15	28	669			
8	1279	12281	16	11	199			

Tabela 2 – Quantidade de lemas e palavras flexionadas presentes no DOE, considerando a quantidade de letras (longitude) da palavra.

Como é possível observar, tanto para os lemas quanto para as palavras flexionadas, há picos, ou seja, a presença de um maior número de palavras para uma determinada longitude. Para os lemas, o pico é atingido com as entradas de 7 letras, enquanto que para as formas, com palavras que possuem 9 letras.

Se representarmos os dados da Tabela 2 na Figura 1, podemos ver que ambas as curvas têm um comportamento tipo Gaussiano: existem valores muito pequenos nas fases inicial e final; tais valores aumentam para a longitude máxima de 7 letras para os lemas e de 9 para as formas e, a partir dessas longitudes, ambas as curvas diminuem novamente.

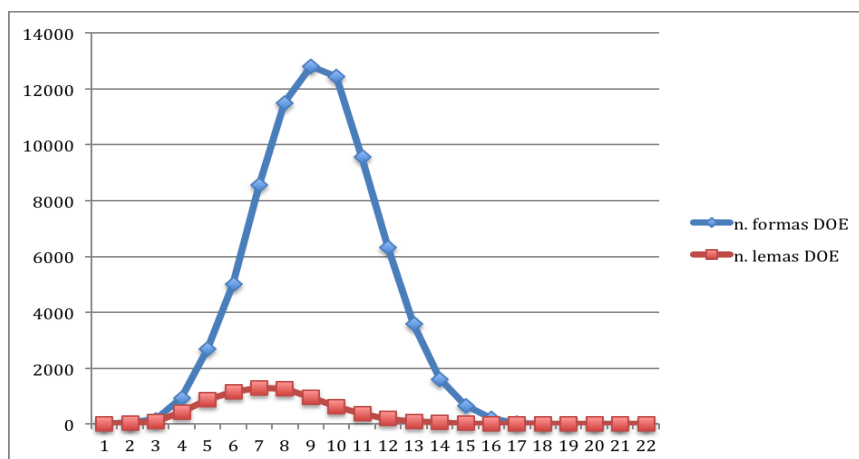


Figura 1– Variação dos lemas e das flexões segundo o número de letras.

A Tabela 3 fornece informações sobre a frequência das letras do alfabeto nos lemas/formas em qualquer posição da palavra. As tabelas 4 e 5 trazem informações relacionadas à distribuição das letras do alfabeto na posição inicial e final da palavra, respectivamente.

<i>letras</i>	<i>lemas</i>	<i>formas</i>	<i>letras</i>	<i>lemas</i>	<i>formas</i>
a/á	7929	116995	Ñ	133	1336
b	988	14262	o/ó	5842	51215
c	3132	31542	P	1585	17648
d	2433	27662	Q	196	1891
e/é	5553	85370	R	5575	73066
f	613	7026	S	2289	73813
g	882	9052	T	3243	31091
h	472	4297	u/ú/ü	1751	17940
i/í	4697	62492	V	645	7238
j	341	3422	W	1	1
k	19	33	X	101	1470
l	2654	22299	Y	116	1211
m	1852	26678	Z	381	4182
n	3470	43036			

Tabela 3– Frequência das letras do alfabeto nos lemas/formas em qualquer posição da palavra.

<i>Posição inicial</i>			
<i>letras</i>	<i>formas</i>	<i>letras</i>	<i>formas</i>
a/á	11320	n	769
b	2120	ñ	5
c	9593	o/ó	1825
d	7651	p	6323
e/é	9035	q	328
f	2516	r	6819
g	1421	s	4331
h	1159	t	3436
i/í	3614	u/ú/ü	395
j	412	v	1710
k	12	x	0
l	1948	y	43
m	3328	z	223

Tabela 4 – Distribuição das letras do alfabeto em posição inicial da palavra.

<i>Posição final</i>			
<i>letras</i>	<i>formas</i>	<i>letras</i>	<i>formas</i>
a/á	10028	ñ	0
b	1	o/ó	7000
c	0	p	1
d	1354	q	0
e/é	7912	r	1502
f	2	s	40186
g	5	t	12
h	4	u/ú/ü	15
i/í	284	v	1
j	1	w	1
k	2	x	0
l	248	y	23
m	2	z	49
n	11703	z	49

Tabela 5– Distribuição das letras do alfabeto em posição final da palavra.

A partir da distribuição das letras do alfabeto na posição final da palavra (Tabela 5), foram selecionadas aquelas com frequência maior que 1.000 (a, d, e, n, o, r, s) e analisadas as sequências constituídas por duas letras. Na Tabela 6, pode-se observar a produtividade de algumas sequências de duas letras.

<i>a</i>	<i>forma</i>	<i>e</i>	<i>forma</i>	<i>o</i>	<i>forma</i>	<i>d</i>	<i>forma</i>	<i>s</i>	<i>forma</i>
ba	1006	ae	8	ao	6	ad	1051	as	10010
ca	286	be	58	bo	65	ed	153	es	5317
da	1601	ce	187	co	335	Id	139	is	11234
ea	75	de	167	do	2885	ud	11	ls	1
fa	11	ee	104	eo	136			os	13601
ga	125	Fe	14	fo	12	<i>n</i>	<i>forma</i>	ps	1
ha	55	ge	17	go	167	an	6253	rs	7
Ia	1845	he	62	ho	78	en	3772	ss	1
ja	85	Ie	95	io	488	in	28	ts	6
ka	1	Je	110	jo	117	on	1641	us	8
la	257	ke	2	lo	286	un	9		
ma	137	Le	226	mo	177				
ña	49	me	94	ño	57	<i>r</i>	<i>forma</i>		
na	309	ñe	37	no	319	ar	991		
oa	2	ne	203	oo	1	er	176		
pa	48	oe	1	po	59	ir	143		
ra	3029	pe	50	ro	475	or	189		
sa	233	Re	2787	so	263	ur	3		
ta	605	se	1377	to	713				
ua	29	Te	1979	uo	24				
va	80	üe	3	vo	106				
ya	31	ue	246	yo	49				
za	128	ve	64	zo	179				
		ye	20						

Tabela 6 – Frequência da sequência de duas letras em palavras flexionadas.

A Tabela 7 mostra as sequências de terminações de 3 letras com uma frequência superior a 100, escolhidas com base na tabela anterior (Tabela 6) com uma frequência maior que 100.

Analizamos as sequências de duas letras, com as informações da Tabela 4, escolhidas dentre aquelas que tiveram frequência superior a 1000, ou seja, igual a 97% do número de formas do DOE. A Tabela 8 mostra a produtividade de alguns grupos de duas sequências de letras iniciais.

<i>A</i>	<i>forma</i>	<i>D</i>	<i>forma</i>	<i>N</i>	<i>forma</i>	<i>o</i>	<i>forma</i>	<i>s</i>	<i>forma</i>	<i>s</i>	<i>forma</i>
Ara	1937	Tad	172	Ran	2627	ndo	1317	mos	10220	nas	310
Ria	1383	dad	143	Ian	1583	ado	1137	ais	4983	cas	284
ada	1145	rad	112	Ren	1378	ido	321	eis	3745	las	253
aba	961			Sen	1291	ero	178	ras	3027	ros	251
era	590	<i>e</i>	<i>forma</i>	Ron	1268	ito	169	éis	2348	sas	231
ida	307	are	1928	Ban	985	nto	167	ias	1833	des	203
Ira	149	ste	1311	Ión	249	ico	164	res	1670	ios	193
ora	144	ase	980	Ten	210	rio	120	das	1602	sos	179
ica	140	ere	452	Tan	208	oso	117	dos	1502	cos	171

cia	136	nte	375	Uen	117	ino	108	ses	1319	nos	164
Ita	133	ese	323	Cen	115	cio	103	bas	1005	ces	162
sta	130	ire	159	Can	109			tas	602	los	144
osa	122	que	146	Nen	108	<i>r</i>	<i>forma</i>	nes	548	mas	141
lla	108					tar	168	tes	411	ues	139
ina	104					dor	112	tos	390	zas	129
						rar	112	les	385	gas	127

Tabela 7 – Frequência de seqüências de três letras em palavras flexionadas.

<i>a/á</i>	<i>forma</i>	<i>c</i>	<i>forma</i>	<i>f</i>	<i>forma</i>	<i>l</i>	<i>forma</i>	<i>p</i>	<i>forma</i>
ab	849	ca	1890	fa	477	la	604	pa	1108
ac	1555	ce	438	fe	222	le	244	pe	1267
ad	790	ch	630	fi	399	li	488	pi	636
ae	12	ci	210	fl	155	ll	372	pl	359
af	412	cl	106	fo	356	lo	103	po	562
ag	773	co	5123	fr	409	lu	137	pr	2127
ah	167	cr	490	fu	498			pu	264
ai	10	cu	706			<i>m</i>	<i>forma</i>		
aj	71			<i>g</i>	<i>forma</i>	ma	1437	<i>r</i>	<i>forma</i>
al	1055	<i>d</i>	<i>forma</i>	ga	318	me	705	ra	468
am	537	da	150	ge	41	mi	253	re	5473
an	666	de	5353	gi	75	mo	662	ri	169
añ	2	di	1545	gl	16	mu	271	ro	602
ap	1240	do	416	go	211			ru	107
aq	10	dr	68	gr	334	<i>o/ó</i>	<i>forma</i>		
ar	939	du	119	gu	426	ob	302	<i>s</i>	<i>forma</i>
as	942					oc	229	sa	983
at	716	<i>e/é</i>	<i>forma</i>	<i>h</i>	<i>forma</i>	od	55	se	853
au	164	eb	2	ha	446	oe	1	sh	6
av	147	ec	67	he	139	of	174	si	421
ay	60	ed	70	hi	172	oh	1	so	1148
az	202	ef	69	ho	206	oi	33	sp	1
		eg	2	hu	196	oj	3	st	1
<i>b</i>	<i>forma</i>	eh	1			ol	176	su	918
ba	746	ej	175	<i>i/í</i>	<i>forma</i>	om	1		
be	261	el	318	ib	5	op	238	<i>t</i>	<i>forma</i>
bi	67	em	1253	ic	6	or	339	ta	474
bl	125	en	3318	id	123	os	67	te	479
bo	396	ep	6	if	1	ot	58	ti	244
br	347	eq	117	ig	19	ou	1	to	516
bu	178	er	171	il	66	ov	12	tr	1583
		es	2011	im	643	ox	1	tu	140
		et	7	in	2672	oy	21		
		eu	6	ir	22	oz	1	<i>v</i>	<i>forma</i>
		ev	224	is	3			va	367
		ex	1216	it	2			ve	439

		ey	1	iz	52			vi	583
								vo	297
								vu	24

Tabela 8 – Frequência da combinação de duas letras iniciais em palavras flexionadas ou formas.

A partir de algumas sequências de duas letras iniciais referentes às formas da Tabela 8, selecionadas entre aquelas com frequência superior a 1.000, foram gerados grupos de três letras no início da palavra. Na Tabela 9, foram incluídos aqueles grupos com frequências superiores a 100.

Nos grupos obtidos a partir da sequência de três letras destacam-se:

- 6 grupos com uma frequência superior a 300 (ver tabela 9): *arr; imp; sal; agr; pla; ama;*
- 39 grupos com uma frequência entre 150 e 300: *ase; for; tor; atr; ven; gua; fun; chi; abo; que; cha; cri; gra; bri; mer; lib; asi; cur; lim; sus; abu; eje; cre; vac; afi; vol; fal; ado; ade; lle; eva; aho; hal; ocu; sac; ata; sen; fre; hab;*
- 72 grupos com uma frequência entre 100 e 150: *cer; sup; bal; pin; san; men; fra; pic; ate; bat; flo; tem; dom; val; agu; cul; jug; pos; cho; gui; uni; bas; fin; sem; adm; bor; ser; llo; ori; ani; lab; qui; cru; aso; pis; ras; ano; ele; equ; aba; lam; ord; sel; mol; ide; tal; adi; agi; tap; abr; obs; sum; azo; ond; sur; afe; ofr; aga; bar; fum; luc; abs; mor; asa; mon; mov; rod; sab; err; pol; her; ara; ten.*

ac/ác	forma	de/dé	forma	ex	forma	re	forma
aca	178	deb	114	Exh	108	rea	117
ace	217	dec	356	Exi	119	reb	163
acl	110	def	178	Exp	529	rec	1091
aco	539	dem	122	Ext	314	red	232
acr	110	den	114			ref	451
acu	193	dep	177	in	forma	reg	282
		der	279	Inc	439	rei	136
al/ál	forma	des	3508	Ind	108	rel	302
alb	119	det	125	Inf	261	rem	125
ale	125	dev	165	Ins	377	ren	202
ali	242			Int	737	rep	473
alm	130	di/dí	forma	Inv	348	res	877
		dif	117			ret	388
ap/áp	forma	dis	777	ma	forma	rev	405
apa	233	div	286	mal	109		
apl	168			man	458	so	forma
apo	270	em/ém	forma	mar	344	sob	402
apr	442	emb	296	mat	145	sol	143

apu	116	emp	780			son	117
				pa	forma	sop	112
ca/cá	forma	en/én	forma	Par	387	sos	114
cal	397	enc	775	Pas	302		
car	247	end	108			tr	forma
cas	195	enf	336	pe	forma	tra	1131
cam	178	eng	270	Pel	157	tri	251
can	177	eno	111	Pen	135	tro	130
cap	140	enr	179	Per	627		
cau	110	ens	283	Pes	134		
		ent	731				
co/có	forma	env	223	pr	forma	Outros grupos	
coa	108			Pre	749	Frec> 300	
col	495	es/és	forma	Pro	1251	arr	631
com	1174	esc	791			imp	525
con	2353	esp	264			sal	370
coo	112	est	809			agr	363
cor	407					pla	333
cos	179					ama	301

Tabela 9 – Frequência de combinações de três letras iniciais em palavras flexionadas.

3. Comparação entre o *Léxico Ativo-Funcional del Escolar Cubano* e o *Diccionario Ortográfico del Español*

Ao analisarmos as palavras do DOE, observamos que estas não incluem algumas formas presentes no LAFEC. Existem 1.804 formas do LAFEC que não foram incluídas entre os lemas do dicionário; a maior parte consiste em formas verbais com enclíticos (terminado em: *el, los, la, las, te, nos, se, me*), alguns tipos e números de substantivos e alguns diminutivos de adjetivos (terminados em: *-ito, -itos, -ita, -itas*).

A entrada do DOE é formada pelo lexema, seguido pelas siglas que indicam seu gênero e número e categoria gramatical. No caso dos verbos, eles sempre aparecem no infinitivo como é lógico, seguidos pela categoria gramatical do verbo e um número escrito entre colchetes que se refere ao seu modelo de conjugação. Entre parênteses, é apresentada a divisão em segmentos para indicar onde se encontra a mudança em relação à norma, ou seja, em relação aos verbos regulares.

Exemplos:

perr -o, -os, -a, -as; sust. m. y f.
volver [38]; v. (v -o -lv -er)

Como pode ser visto na Tabela 10, o número de formas do DOE (80.336) é muito maior do que as formas do LAFEC (17.967); no entanto, as entradas no DOE totalizaram apenas 7.535.

Além disso, na tabela 10, mostra-se a distribuição do número de palavras flexionadas diferentemente, tanto para o DOE quanto para o LAFEC, ao variar a longitude (número de letras); o número de lemas para o DOE também é apresentado na mesma tabela quando a longitude varia.

Nas três últimas colunas da Tabela 10, foram acrescentados alguns dados mais significativos que mostram o número de formas diferentes no LAFEC, tanto em relação à longitude da palavra quanto à frequência no léxico, ou seja, sua ocorrência na produção dos estudantes. Para sermos mais precisos, as informações, nas três últimas colunas, referem-se ao léxico com frequência ≥ 3 , ≥ 5 e ≥ 10 entre palavras de três a catorze letras.

A figura 2 indica o número de lemas e palavras flexionadas referentes ao DOE e o número de palavras flexionadas para o LAFEC, de acordo com a longitude das palavras.

Long.	Formas DOE	Lemas DOE	Formas LAFEC	Formas LAFEC (f. ≥ 3)	Formas LAFEC (f. ≥ 5)	Formas LAFEC (f. ≥ 10)
1	5	5	5			
2	48	34	45			
3	184	95	160	130	117	102
4	1085	431	789	564	468	360
5	3041	884	1732	1135	922	647
6	5619	1150	2418	1335	983	641
7	9344	1301	3125	1482	1060	654
8	12281	1279	3009	1306	897	512
9	13505	958	2518	935	617	309
10	12858	630	1941	658	404	201
11	9781	371	1119	348	220	101
12	6423	196	600	167	96	47
13	3617	94	275	72	40	15
14	1623	61	129	35	20	8
15	669	28	63			
16	199	11	22			
17	41	3	8			
18	8	3	5			
19	1	0	1			
20	0	0	0			
21	2	1	1			
22	2	0	2			

Tabela 10 – Número de lemas/formas no DOE e número de formas no LAFEC para diferentes longitudes.

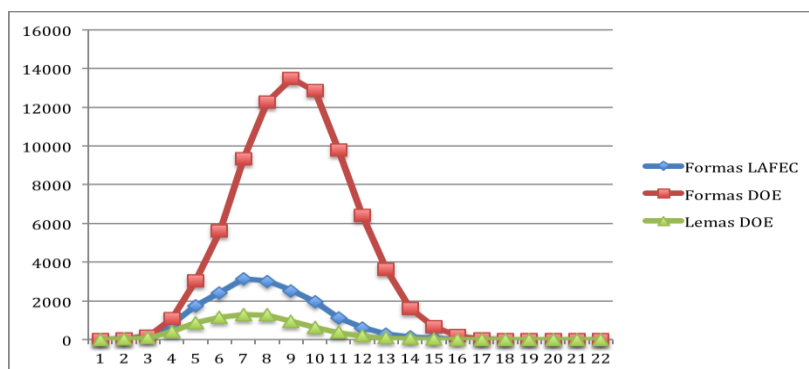


Figura 2 – Número de lemas e palavras flexionadas (DOE) e número de palavras flexionadas (LAFEC), de acordo com o número de letras das palavras.

A figura 2 mostra que o número de formas diferentes do LAFEC e do DOE é variável. No caso do LAFEC, 99% das palavras contêm entre três e catorze letras e, no DOE, o índice é de 98% nessa faixa de longitude.

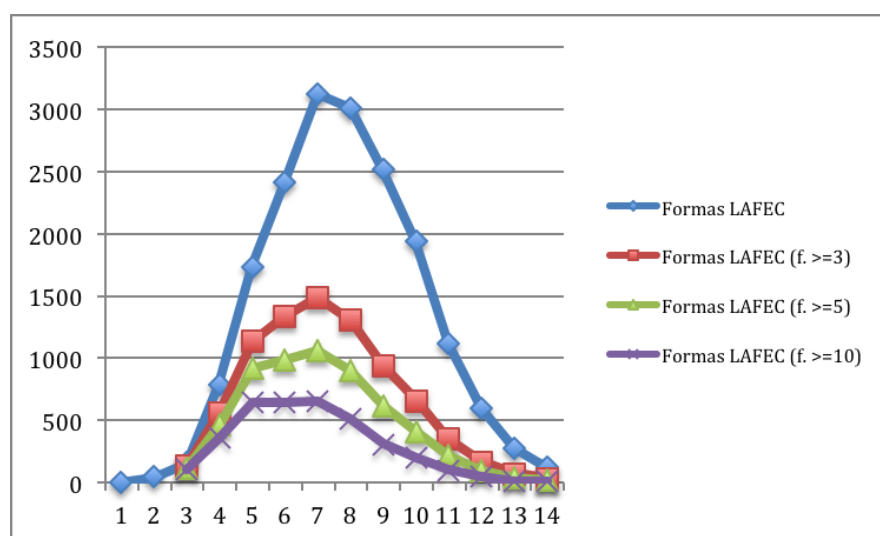


Figura 3 – Número de palavras flexionadas de acordo com o número de letras das palavras e frequências (LAFEC).

A Figura 3 mostra o número de palavras flexionadas no LAFEC, considerando a longitude das palavras entre três e catorze letras, para frequências $\Rightarrow 3$, $\Rightarrow 5$, $\Rightarrow 10$. Ao analisar a Figura 3, fica evidente que, para o LAFEC, as palavras mais utilizadas pelas crianças são aquelas com 7 letras, enquanto que para as palavras geradas com DOE (Fig. 2), o pico ocorre nas palavras de 9 letras. Não é difícil entender que a obtenção de palavras do DOE é mais produtiva do que a dos alunos do ensino fundamental.

Considerando a decisão de selecionar formas com a longitude variável entre 3 e 14 caracteres pertencentes ao LAFEC, um subconjunto de palavras com uma extensão entre 3 e 14 letras (LAFEC-P) foi extraído e, em seguida, um subconjunto de palavras com uma frequência $\Rightarrow 3$ (LAFEC-PF3).

Há quatro conjuntos de dados que usaremos nesta seção dedicada à comparação entre o LAFEC e DOE: as formas geradas com o DOE, a informação do LAFEC e seus subconjuntos (LAFEC-P e LAFEC-PF3). Com esses grupos, combinações possíveis de 2, 3 e 4 letras foram geradas. Esses resultados foram comparados entre si para tentar identificar as sequências relevantes de 2, 3 e 4 letras a partir do menor subconjunto (LAFEC-PF3), a fim de verificar se essas sequências eram relevantes também para as do DOE, ou seja, para o maior conjunto.

A Figura 4 mostra os valores das diferentes sequências de duas letras nos diferentes conjuntos e para as sequências com frequência $\Rightarrow 3$, isto é, para sequências que aparecem em três ou mais palavras diferentes. Como é possível observar, o número de diferentes sequências nos conjuntos varia pouco, oscila entre 344 sequências para a DOE, 355 para o LAFEC e 329 para o LAFEC-P, chegando a 291 para o LAFEC-PF3, o menor conjunto.

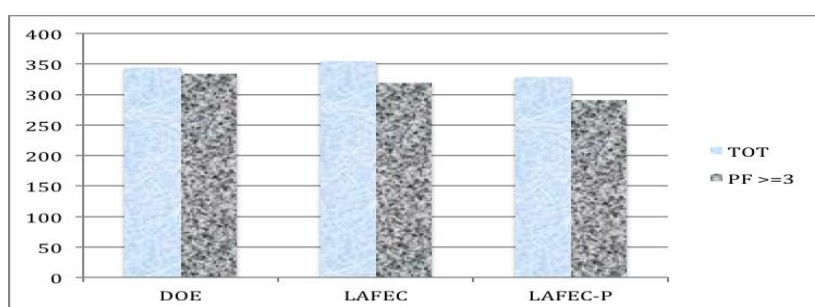


Figura 4– Valores das diferentes sequências de duas letras nos diversos conjuntos.

Quando as combinações de duas letras do DOE e do LAFEC-PF3 são comparadas, obtém-se 344 combinações para o DOE, entre as quais 21 não aparecem no conjunto do LAFEC-PF3. Essas combinações são: *bn*: 57, *dj*: 2, *dv*: 54, *dy*: 54, *ih*: 2, *ik*: 8, *ix*: 54, *iy*: 2, *jt*: 2, *lh*: 2, *nh*: 58, *nn*: 2, *ow*: 1, *pc*: 8, *ps*: 3, *ry*: 1, *sk*: 2, *uq*: 12, *xh*: 108, *xq*: 4, *zn*: 2⁹.

Das 21 combinações de duas letras, analisaremos as seis que possuem frequência superior a 50 em diferentes palavras: a sequência *xh* que aparece em 108 palavras diferentes, refere-se às conjugações dos verbos *exhibir* e *exhortar*, cujo uso no LAFEC teve uma frequência muito baixa. O mesmo acontece com a sequência *nh* que se refere às formas do verbo *anhelar* (ansiar¹⁰) e às formas do adjetivo *inhumano* (desumano); a combinação *ix* refere-se às formas do verbo *asfixiar*; o *dy* ao verbo *coadyuvar* (auxiliar); *dv* refere-se ao verbo *advertir* a combinação *bn*, ao verbo *abnegar*.

Ao analisar o conjunto LAFEC-P, 329 combinações de duas letras foram obtidas. Apenas 5 combinações do LAFEC-P não estão presentes no DOE: *ek*, *kw*, *wo*, *rk*, *yi*. As combinações *ek*, *kw*, *wo*, *rk* aparecem em uma única palavra. A baixa frequência é lógica

⁹ Os números representam a frequência da combinação de duas letras em diferentes palavras.

¹⁰ N.T.: Foram indicados os equivalentes em português somente para as palavras cujo significado era menos transparente.

devido ao fato de serem estrangeirismos pouco utilizados. No caso de *ek*, *kw* e *wo*, as combinações referem-se à palavra *taekwondo* e, no caso de *rk*, *asnorkel*.

Para a combinação *yi*, a sequência foi encontrada em palavras diferentes, tal como nos diminutivos *cayito*, *rayito*, *hoyito*, *rayita*, *hoyitos*, *playita*, *rayitas*, *rayitos*, *arroyito*¹¹.

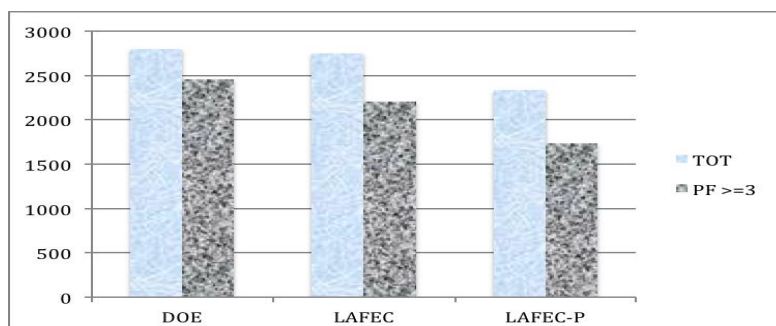


Figura 5 – Valores das diferentes sequências de três letras nos diferentes conjuntos.

Em relação à sequência de três letras, na Figura 5, pode-se observar que há mudanças significativas nos três conjuntos (DOE, LAFEC e LAFEC-P). Para o DOE existem 2.800 sequências diferentes, 2.747 para o LAFEC e 2.333 para o LAFEC-P. As diferentes sequências para o conjunto LAFEC-PF3 totalizaram 1.735 (sequências envolvendo três ou mais palavras diferentes).

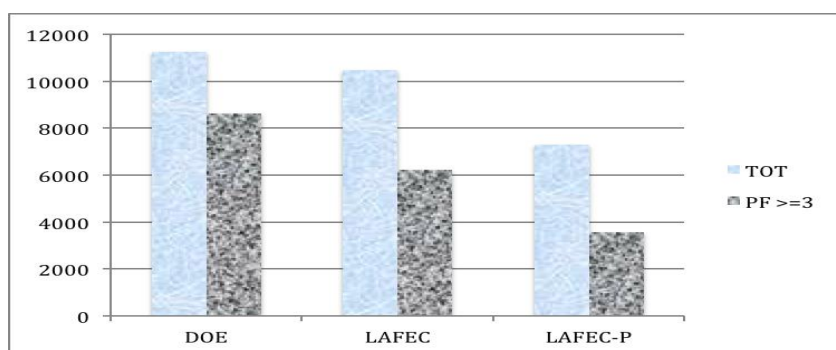


Figura 6 – Valores das diferentes sequências de quatro letras nos diversos conjuntos

A figura 6 mostra a sequência de quatro letras. As variações são maiores nos três conjuntos. Para o DOE existem 11.245 sequências diferentes, 10.478 para o LAFEC e 7.295 para o LAFEC-P. As diferentes sequências para o conjunto LAFEC-PF3 totalizaram 3.567 (sequências envolvendo três ou mais palavras diferentes).

¹¹ N.T.: ilhazinha, raiozinho, burquinho, linhazinha, buraquinhos, praiazinhas, linhazinhas, raiozinhos, riozinho.

Com base nos cálculos feitos, percebe-se que muitas sequências não são comuns aos três conjuntos; algumas pertencem apenas ao DOE, outras apenas ao LAFEC-P e outras ao LAFEC-PF3. Por exemplo, se consideramos um conjunto de 4 letras que pertencem apenas ao LAFEC-PF3, onde a frequência das sequências aparece em pelo menos dez palavras, obteremos 32 grupos (ver tabela 11).

Depreende-se da Tabela 11 que esses grupos de 4 letras estão presentes principalmente na parte interna das palavras; apenas quatro sequências encontram-se no final das palavras (*irse, selo, erlo, rlos*), três na parte interna e final (*ogía, selo, erlo*) e três no início e no meio da palavra (*tiri, irno, darl*).

<i>total</i>	<i>princípio</i>	<i>interior</i>	<i>final</i>	<i>sequência</i>
10	0	10	0	derl
10	0	10	0	eint
10	0	10	0	iarl
10	0	10	0	jarl
10	0	10	0	zarl
11	0	11	0	gerl
11	0	11	0	ueli
11	1	10	0	tiri
12	0	12	0	sarl
13	0	13	0	cerl
13	0	13	0	larl
13	0	13	0	varl
13	0	13	0	rarm
14	0	5	9	ogía
14	0	14	0	rarn
14	0	14	0	tars
15	0	15	0	nerl
16	0	16	0	narl
17	0	0	16	irse
18	0	8	10	selo
19	0	19	0	rars
20	0	20	0	tarn
21	1	20	0	irno
23	2	21	0	darl
26	0	26	0	ndom
27	0	27	0	carl
28	0	28	0	rsel
41	0	18	23	erlo
42	0	42	0	rarl
48	0	48	0	tarl
62	0	0	62	rlos
87	0	87	0	arno

Tabela 11 – Grupo de 4 letras pertencentes ao LAFEC-PF3 e com frequência de sequências igual ou superior a 10.

Nesse caso, por exemplo, a sequência de quatro letras **arno**, com frequência 87 em posição interna da palavra, corresponde às formas: *ayudarnos, buscarnos, darnos, saludarnos, etc.*, ou seja, todas as formas verbais com enclíticos. Inclusive, nos casos da combinação **rlos**, com frequência 62 no final da palavra, também estamos diante de formas verbais com enclíticos: *asarlos, curarlos, dejarlos, sacarlos, etc.*; no caso de **erlo**, refere-se a: *conocerlo, esconderlo, protegerlo, etc.*; no caso de **selo** a: *comérselo, entregárselo, llevárselo, etc.*

Para a combinação **ueli**, no interior da palavra e com frequência 11, temos a presença de diminutivos como *abuelito, abuelita, escuela, vuelitos*¹². Com frequência inferior a 10, destaca-se o uso de diminutivos com as seguintes sequências **brit**: *librito*, **ayit**: *rayito*, **ceci**: *lucecita*, **jari**: *pajarito*, **reri**: *sombrerito*, **añit**: *montañitas*, **yito**: *arroyito, hoyitos*¹³. Todos os casos acima estão em correspondência correta com o vocabulário dos estudantes de nível de ensino analisado. Exemplos de palavras que terminam em **ogía** são: *ideología, mitología, arqueología, ginecología, etc.*

No grupo das 32 sequências, há algumas, como **erlo, rlos, darl, arno, rarn, nerl**, que estão na mesma palavra, como por exemplo, *ponerlos*, constituída de duas sequências no interior da palavra (**nerl, erlo**) e uma no final (**rlos**).

Analisaremos, abaixo, um grupo de sequências de quatro letras pertencentes ao DOE. Trata-se de um pequeno subconjunto em que as sequências têm uma frequência igual ou superior a 100, ou seja, aparecem em pelo menos cem palavras. Os grupos pertencentes ao DOE, com frequência igual ou superior a 100, correspondem a um total de 1.145 sequências, das quais trinta apenas se encontram no DOE (ver Tab.12) e não nos demais conjuntos (LAFEC, LAFEC-P e LAFEC-PF3).

<i>total</i>	<i>príncipi o</i>	<i>Interior</i>	<i>final</i>	<i>sequência</i>
100	0	0	100	néis
102	0	102	0	east
105	0	0	105	céis
107	0	0	107	cáis
108	0	0	108	uéis
116	0	87	29	jase
118	0	118	0	zast
145	0	145	0	jarí
155	0	155	0	tirí
163	0	123	40	mase
168	0	126	42	iase
180	0	135	45	sase
196	0	147	49	dase

¹² N.T.: vovozinho, vovozinha, escolinha, voozinho.

¹³ N.T.: livrinho, raiozinho, luzinha passarinho, chapeuzinho, montanhazinhas, riozinho, buraquinho.

202	0	0	202	táis
204	0	153	51	ease
236	0	177	59	zase
294	0	294	0	esei
295	1	294	0	erai
295	0	295	0	erei
320	0	240	80	nase
952	0	952	0	asei
954	0	954	0	arai
954	2	952	0	arei
955	0	955	0	abai
956	0	0	956	bais
1246	0	0	1246	teis
1247	0	0	1247	rais
1247	0	0	1247	reis
1267	0	1267	0	ría
1537	0	0	1537	íais

Tabela 12 – Grupo de sequências de quatro letras com frequência igual ou superior a 100 (DOE).

Focalizaremos nas sequências de terminações *iais*, *reis*, *rais*, *teis*, *bais*, *táis*, *uis*, *cáis*, *céis*, *néis*. No caso de *íais*, presente em 1.537 palavras diferentes, a sequência refere-se à forma verbal do futuro do pretérito do indicativo da segunda pessoa do plural (vosotros/vosotras); para *reis*, as palavras correspondem à forma verbal do futuro do subjuntivo na segunda pessoa do plural; para *teis*, à forma verbal do pretérito perfeito do indicativo na segunda pessoa do plural; para *bais*, à forma verbal do pretérito imperfeito do indicativo na segunda pessoa do plural; para *táis*, *cáis*, *céis* às formas verbais do presente do indicativo na segunda pessoa do plural; para *uéis*, à forma verbal do presente do subjuntivo na segunda pessoa do plural e para *néis*, à forma verbal do presente do indicativo na segunda pessoa do plural.

Como se sabe, na América Hispânica, a segunda pessoa do plural *vosotros/vosotras* não é usada, razão pela qual, evidentemente, as formas verbais mencionadas anteriormente da segunda pessoa do plural não foram utilizadas nos escritos – e muito menos nas exposições orais – das crianças cubanas. A análise comparativa do DOE e do LAFEC produziu uma valiosa informação sobre o uso de palavras pelos estudantes. Os resultados obtidos relacionados às diferentes sequências são uma fonte rica de dados úteis para estudos de reconhecimento de fala e estudos de escrita de textos, bem como para pesquisas psicolinguísticas e para a preparação de exercícios didáticos dirigidos aos estudantes.

4. Os anagramas

Estudos sobre anagramas são frequentemente muito úteis na educação e reabilitação. A partir de algumas pesquisas neurolinguísticas¹⁴, sabe-se que as informações sobre as sequências de letras são muito importantes para o estudo dos problemas apresentados por sujeitos disgráficos e afásicos. Com a extensa base de dados do DOE foram gerados todos os códigos alfabéticos ou alfa-code (a-code¹⁵) tanto para os lemas quanto para as formas.

Destaca-se que alguns a-code representam apenas uma palavra, outros podem gerar mais de uma. Foram calculados todos os possíveis anagramas com seu a-code¹⁶ que podem gerar mais de uma solução na língua real, ou seja, palavras que existem no espanhol. No que se refere aos lemas, o número dos possíveis a-code, no DOE, chega a 218 com uma produção de 457 anagramas. Para as formas, o número dos possíveis a-code atinge 4.541, e o número dos anagramas chega a 10.182.

Nas tabelas 13 e 14, são apresentados o número de lemas e formas e as porcentagens de palavras com e sem anagramas, subdivididos de acordo com a quantidade de letras.

nr. let.	nr. lem.	% semanag.	% anag.	nr. let.	nr. lem.	% sem. anag.	% anag.
1	5	100,0	0,0	12	196	100,0	0,0
2	34	82,4	17,6	13	94	100,0	0,0
3	95	87,4	12,6	14	61	100,0	0,0
4	431	78,4	21,6	15	28	100,0	0,0
5	884	86,0	14,0	16	11	100,0	0,0
6	1150	92,3	7,7	17	3	100,0	0,0
7	1301	94,7	5,3	18	3	100,0	0,0
8	1279	96,8	3,2	19	0	-	-
9	958	98,1	1,9	20	0	-	-
10	630	99,0	1,0	21	1	100,0	0,0
11	371	100,0	0,0	22	0	-	-

Tabela 13 – Porcentagem de lemas com e sem anagramas de acordo com a quantidade de letras.

nr. let.	nr. form.	% semanag.	% anag.	nr. let.	nr. form.	% sem. anag.	% anag.
1	5	100,0	0,0	12	6423	97,9	2,1
2	48	75,0	25,0	13	3617	98,5	1,5
3	184	82,1	17,9	14	1623	99,6	0,4
4	1085	68,2	31,8	15	669	100,0	0,0

¹⁴ As análises estatísticas realizadas em Deloche, Debili, Andreewsky (1980) referem-se às listas de palavras inglesas e francesas extraídas de um *corpus* de referência e de listas de frequências.

¹⁵ Sequências de letras em ordem alfabética de uma palavra.

¹⁶ Por exemplo, o a-code **aglo** pode gerar os anagramas: **algo**, **lago**.

5	3041	73,0	27,0	16	199	100,0	0,0
6	5619	77,7	22,3	17	41	100,0	0,0
7	9344	76,7	23,3	18	8	100,0	0,0
8	12281	83,6	16,4	19	1	100,0	0,0
9	13505	86,5	13,5	20	0	-	-
10	12858	92,3	7,7	21	2	100,0	0,0
11	9781	94,7	5,3	22	2	100,0	0,0

Tabela 14 – Porcentagem de formas com e sem anagramas de acordo com a quantidade de letras.

Nas duas tabelas seguintes, podemos ver os cálculos em relação ao número de letras nos lemas (Tabela 15) e nas formas (Tabela 16). Da distribuição dos a-code podem-se originar 1 palavra, 2 palavras, 3 palavras, etc.

<i>l. lemas</i>	<i>n. lemas</i>	<i>a-code 1</i>	<i>a-code 2</i>	<i>a-code 3</i>	<i>a-code 4</i>
1	5	5	0	0	0
2	34	28	3	0	0
3	95	83	6	0	0
4	431	338	36	7	0
5	884	760	49	6	2
6	1150	1062	41	2	0
7	1301	1232	33	1	0
8	1279	1238	19	1	0
9	958	940	9	0	0
10	630	624	3	0	0
11	371	371	0	0	0
12	196	196	0	0	0
13	94	94	0	0	0
14	61	61	0	0	0
15	28	28	0	0	0
16	11	11	0	0	0
17	3	3	0	0	0
18	3	3	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	1	1	0	0	0
22	0	0	0	0	0

Tabela 15 – Distribuição dos a-code nos lemas de acordo com a quantidade de letras.

<i>l. formas</i>	<i>n. formas</i>	<i>a-code 1</i>	<i>a-code 2</i>	<i>a-code 3</i>	<i>a-code 4</i>	<i>a-code 5</i>	<i>a-code 6</i>	<i>a-code 7</i>	<i>a-code 8</i>
1	5	5	0	0	0	0	0	0	0
2	48	36	6	0	0	0	0	0	0
3	184	151	15	1	0	0	0	0	0

4	1085	740	107	34	6	1	0	0	0
5	3041	2220	274	57	15	6	2	0	0
6	5619	4365	428	84	19	9	3	1	0
7	9344	7163	691	159	46	11	8	5	0
8	12281	10268	774	107	25	6	1	0	1
9	13505	11681	687	105	25	7	0	0	0
10	12858	11870	427	36	4	2	0	0	0
11	9781	9264	236	12	1	1	0	0	0
12	6423	6289	64	2	0	0	0	0	0
13	3617	3563	27	0	0	0	0	0	0
14	1623	1617	3	0	0	0	0	0	0
15	669	669	0	0	0	0	0	0	0
16	199	199	0	0	0	0	0	0	0
17	41	41	0	0	0	0	0	0	0
18	8	8	0	0	0	0	0	0	0
19	1	1	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
21	2	2	0	0	0	0	0	0	0
22	2	2	0	0	0	0	0	0	0

Tabela 16 – Distribuição dos a-code nas formas de acordo com a quantidade de letras.

As duas tabelas anteriores mostram que, no DOE, a maioria dos a-code geram palavras únicas, tanto para os lemas como para as formas. Nos lemas, a porcentagem de palavras únicas é de 93,9% em comparação com 6,1% nos anagramas; nas formas, o percentual de palavras únicas é de 87,3%, em contraste com 12,7% nos anagramas.

A concentração do maior número de anagramas, tanto para os lemas como para as formas, encontra-se no a-code= 2, ou seja, quando um a-code origina duas palavras diferentes (anagramas).

Para os lemas, a concentração do maior número de anagramas encontra-se nas palavras de 4 a 7 letras e, para as formas, em palavras com uma longitude entre 5 e 10 letras.

É interessante notar que, para o a-code = 8 *acenorst*, foram obtidos os seguintes 8 anagramas: *canteros* (canteiros), *cartones* (papelões), *contares* (segunda pessoa do singular do verbo *contar* no futuro do subjuntivo), *contraes* (segunda pessoa do singular do verbo *contraer* no presente do indicativo), *cornetas*, *cortasen* (terceira pessoa do plural do verbo *cortar* no pretérito imperfeito do subjuntivo), *costaren* (terceira pessoa do plural do verbo *costar* no futuro do subjuntivo).

Estudos futuros sobre a variante do espanhol cubano – por exemplo, o projeto em curso *Corpus del español de Cuba* (CECu, 1995-2014), realizado entre o Centro de Linguística Aplicada, a Empresa de Desarrollo de Aplicaciones, Tecnología y Sistemas (DATYS), ambos de Santiago de Cuba, e o Instituto de Literatura e Linguística "José Antonio Portuondo", de Havana –, poderiam ser comparados com os resultados alcançados com o DOE para confirmar as características na formação de anagramas, ou

seja, se há um aumento de sua porcentagem, ao aumentar o número de formas, e se a concentração do maior número de anagramas permanece no a-code= 2.

5. Conclusões

Neste artigo, foi feita uma análise detalhada das características de um subconjunto do espanhol de Cuba com base no LAFEC e no DOE. O estudo de lemas, palavras flexionadas, sequências de letras e extensão de palavras, com o auxílio de tabelas, figuras e estatísticas, ofereceu uma visão mais completa das características distributivas da variante cubana do espanhol presente nos estudantes cubanos.

Finalmente, gostaríamos de indicar os endereços da Internet para *download* dos seguintes estudos:

- *Diccionario Ortográfico del Español*
<http://www.ge.ilc.cnr.it/page.php?ID=dic-cubano&lingua=it>
- *Vocabulario Inverso y Anagramas del Español*
<http://www.ge.ilc.cnr.it/page.php?ID=inverso&lingua=it>

Referências

DELOCHE, G., DEBILI, F., ANDREEWSKY, E. Order Information Redundancy of Verbal Codes in French and English: Neurolinguistic Implications, *Journal of Verbal Learning and Verbal Behavior*, Nr. 19, 1980, p. 525 – 530.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C. Rimas y anagramas en el léxico del escolar cubano. Actas-I, VII Simposio Internacional de Comunicación Social, Santiago de Cuba 22-26 de enero 2001. Ediciones Centro de Lingüística Aplicada y Universidad de Málaga, 2001, p. 101-105.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C., RUIZ MIYARES, L. *Diccionario Ortográfico del Español basado en el léxico del escolar cubano*, CNR – Istituto per i Circuito Elettronici, Genova, Italia, Centro de Lingüística Aplicada, Santiago de Cuba, Cuba. Genova: Ed. CNR – Istituto per i Circuito Elettronici, 2000a.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C., RUIZ MIYARES, L. *Vocabulario Inverso y Anagramas del Español basado en el léxico del escolar cubano*; CNR – Istituto per i circuito Elettronici, Genova, Italia, Centro de Lingüística Aplicada, Santiago de Cuba, Cuba; Genova: Ed. CNR – Istituto per i Circuito Elettronici, 2000b.

MIYARES BERMÚDEZ, E. (DIR.). *Léxico Activo-Funcional del Escolar Cubano*, Santiago de Cuba: Ediciones Centro de Lingüística Aplicada, 2006.

MIYARES BERMÚDEZ, E. (DIR.). *Diccionario básico escolar*. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada y Editorial Oriente, 2014.

RATTI, D., MARCONI, L., BURANI, C. *Radiografia dell'italiano*, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma, *La Sapienza*, N. 2, 1995.

RATTI, D., MARCONI, L., MORGAVI, G., ROLANDO, C. *Flessioni-Rime-Anagrammi*. Bolonha: Ed. Zanichelli, 1988.

RUIZ HERNÁNDEZ, V., MIYARES BERMÚDEZ, E. Aplicaciones metodológicas para elevar la cultura del idioma. In: *Estudios de Comunicación Social, 25 años en defensa del idioma*. La Habana: Editorial Academia, 1997, p. 66–73.

ZINGARELLI, N., ZANICHELLI. *Il nuovo Zingarelli minore*. Bologna: Zanichelli, 1987.

Como citar este texto (ABNT):

MARCONI, L.; MIYARES, L.R.; CUTUGNO, P. Tradução de Cyrano da Rosa Silva. Características distributivas do Espanhol de Cuba a partir da análise de dois estudos linguísticos. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 126-148, 2018.

OS VERBOS DE CONSTATAÇÃO NO TEXTO CIENTÍFICO¹

Francis Grossmann²

Tradução: Alex de Cássio da Silva³

Supervisão e revisão: Patrícia Reuillard⁴

Revisão técnica: Sandra Loguercio⁵

1. Introdução

A maneira mais simples de expressar uma constatação é a afirmação, cuja forma canônica é a frase simples:

(1) *La terre tourne autour du soleil.*

[A Terra gira ao redor do sol.]

Uma frase como:

(2) *Je constate que la terre tourne autour du soleil.*

[Eu constato que a Terra gira ao redor do sol.]

não modifica o valor de verdade do enunciado. Por outro lado, sinaliza uma admissão da parte do enunciador e indica que a informação foi obtida por ele através de uma base empírica. Ademais, esse tipo de enunciado é frequentemente formulado como resultado de um raciocínio inferencial:

(3) *Pierre est encore en retard. Je constate qu'il a bien du mal à se réveiller.*

[Pedro está atrasado novamente. Eu constato que ele tem bastante dificuldade em acordar.]

¹ A presente tradução foi autorizada para ser publicada em português, nos Cadernos de Tradução do IL (UFRGS), pela Editora Presses Universitaires de Rennes (PUR) e pelo autor Francis Grossmann. Referência bibliográfica completa do artigo original: GROSSMANN, Francis. "Les verbes de constat dans l'écrit scientifique". In: TUTIN, A.; GROSSMANN, F. (org.) *L'écrit scientifique: du lexique au discours*. Rennes: PUR, p. 85-100, 2013.

² Laboratoire de Linguistique et de Didactique du Français Langue Étrangère et Maternelle (LIDILEM), Université Grenoble 3 – Stendhal.

³ Bacharelando em Letras – Tradução Português/Francês, UFRGS.

⁴ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁵ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

Está em cena aqui a própria operação de constatação, que permite ao enunciador explicar suas conclusões, mas também, às vezes, em função do contexto, marcar seu ponto de vista (Pedro é preguiçoso, ele não é sério, etc.). Comparados aos verbos de julgamento, os verbos marcadores de constatação partem da observação, daí a ancoragem no semantismo da percepção que caracteriza muitos deles, sem que, no entanto, toda constatação precise necessariamente de uma operação perceptiva⁶.

Propomos aqui examinar o funcionamento dessa classe, ou melhor, de uma parte dela, analisando o papel das unidades que a constituem nos gêneros específicos de escrita científica. Os verbos de constatação estão longe de serem o único meio de explicitar uma constatação científica, além do mais, eles têm outras funções além dessa. Nossa abordagem pretende ser essencialmente lexical, levando em conta as construções sintáticas priorizadas e algumas características semânticas, como o aspecto. Por outro lado, tentaremos especificar, na medida do possível, como os verbos estudados, para além de um funcionamento semelhante, desempenham papéis específicos no discurso científico de acordo com seu semantismo. Trabalhamos, portanto, com duas hipóteses: por um lado, acreditamos que há de fato uma classe de verbos constativos, que globalmente têm um funcionamento comum nos gêneros científicos e que se distinguem dos verbos com função metatextual, mesmo que estes frequentemente recorram ao mesmo material lexical; por outro lado, esses verbos funcionam textualmente de acordo com o *status* epistêmico. Assim, verbos como *noter*, *voir*, *remarquer* e *constater* [observar, ver, observar, constatar], além de seu parentesco funcional, teriam uma especialização epistêmica condicionada por seu perfil semântico e sintático que explicaria seu funcionamento específico no âmbito do discurso científico⁷.

Nossa análise apoia-se em uma exploração do *corpus* de língua francesa do SCIENTEXT⁸, que inclui todos os gêneros (com exceção dos pareceres) e todas as disciplinas, consistindo em cerca de 4,8 milhões de palavras. A discussão que propomos não está centrada nas diferenças observadas em função dos gêneros ou das disciplinas, mas busca lançar luz sobre algumas das principais tendências no que diz respeito ao uso dos verbos estudados na totalidade do *corpus*.

⁶ Delplanque (2006) ressalta o fato de que esses verbos, como os verbos de julgamento (*Je trouve que...* [Eu acho que...]), são explicitamente construídos a partir de um sujeito enunciador referindo-se a alguém do mundo real. Essa característica opõe-se à dos verbos de aparência em que o enunciador desocupa o lugar de "construtor", mesmo que reaparecendo, em certos casos e sob certas condições, como "receptor" do processo: *Pierre m'a l'air triste* [Pierre me parece triste].

⁷ Damos continuidade aqui a uma reflexão iniciada com Agnès Tutin (ver GROSSMANN & TUTIN, 2010a e 2010b).

⁸ N.T.: trata-se de um projeto interuniversitário – reunindo os laboratórios LIDILEM, da Universidade Grenoble 3-Stendhal, Littérature Langage Société, da Universidade de Chambéry, e LiCorn, da Universidade de Bretagne Sud – que, entre outros recursos, constituiu *corpora* de textos científicos em inglês e francês disponíveis *online* e que abrangem diversas áreas do conhecimento, servindo de ferramenta para estudos de forma, estilo, colocações etc.

2. Características dos verbos de constatação

2.1. Definição

Chamamos convencionalmente de verbos de constatação (ou constativos⁹) os verbos sinônimos de *constater*, utilizados em nosso *corpus* de escrita científica, que desempenham pragmaticamente o papel de marcador de constatação, no sentido definido na introdução. O *Trésor de la langue française informatisé* (TLFi) apresenta dois empregos principais, deixando de lado a acepção jurídica ("atestar por um ato oficial"): a) no emprego imperfectivo, *constater* é "estabelecer a existência ou autenticidade de um fato após análise"; na atividade científica, é, mais particularmente, "assegurar-se, por meio de observações científicas, da realidade de um fato"; b) no emprego perfectivo, trata-se de "tomar conhecimento de algo", ou, em uma construção mais interessante para nós aqui, "dar-se conta da existência de um fato"¹⁰. Como se vê, a polissemia é bem pequena para esse verbo. Porém, o funcionamento desses verbos pode ser bem diferente na escrita científica: em um sentido "textual", a constatação enunciada corresponde a pontos (fatos, ideias, hipóteses...) considerados indiscutíveis em trabalhos anteriores, ou que serão demonstrados no decorrer do texto. A estrutura prototípica aqui é a construção com *comme* comportando um verbo no *passée composé*¹¹ [passado composto] com valor de concluído, ou no *futur simple*¹² [futuro simples] ou perifrástico: *comme nous avons pu le voir (le constater, etc.)* [como nós pudemos ver (constatar etc.), *comme nous allons le voir, on le verra...* etc.¹³ [como nós veremos, se verá... etc.]. A característica dessas construções é, nesse caso, ter uma função metatextual e/ou evidencial (Grossmann e Tutin, 2010a; 2010b): o autor mostra que o fato que ele informa foi constatado (fundamentado e comprovado) ou será ainda constatado; a estrutura desempenha, essencialmente, um papel retórico de reforçar a argumentação e/ou de orientar o leitor no desenvolvimento do texto. Nesse emprego, o único uso possível é o do *nous* [nós] inclusivo, estando a utilização do *je* [eu] totalmente descartada (Grossmann e Tutin, 2010a; Tutin 2010a):

(4)?*comme je l'ai vu sur la figure X...*
[? como eu vi na figura X...]

⁹ Em um sentido, obviamente, bem distinto daquele dado por AUSTIN, em *How to do Things with Words* (1962).

¹⁰ Uma citação fornecida pelo TLFi oferece um exemplo literário desse emprego: "No hotel onde me encontro, onde ouço falar uma porção de estrangeiros, constato novamente que a língua francesa tem um fluxo qual nenhuma outra". André GIDE, *Journal*, p.1202.

¹¹ N.T.: corresponde ao pretérito perfeito no português.

¹² N.T.: correspondente ao futuro do presente em português.

¹³ DEBAISIEUX e MARTIN (2010, p.321) lembram que "as estruturas com *comme* majoritariamente introduzem "comparativas metaenunciativas" (AUTHIER-REVUZ, 1995, p.115). Este ponto também é relatado por CAPPEAU & SAVELLI (2001).

Deixando de lado este tipo particular de estrutura, nosso interesse se centrará em um outro tipo de enunciado, que nomearemos constatação assertiva: o autor faz a constatação no próprio enunciado, com o auxílio de uma estrutura verbal no presente (tipicamente uma completiva¹⁴: *nous constatons que X* [nós constatamos que X]). Esse tipo de constatação pode ser confirmado por um elemento de prova (figura, esquema, exemplo) figurando no cotexto próximo, embora nem sempre seja esse o caso. Acreditamos que as constatações assertivas têm uma função mais central para a exposição dos resultados do que as estruturas metaenunciativas com *comme* [como], na medida em que permitem evidenciar o resultado de uma demonstração ou da observação de elementos empíricos, e que seria então mais interessante identificá-las, por exemplo, do ponto de vista da busca de informações. O teste do *je* fornece um resultado diferente daquele observado para as estruturas metaenunciativas: o *je* do autor-pesquisador, embora pouco representado¹⁵, permanece teoricamente sempre possível¹⁶ – cf. o exemplo construído em (5) – o que justifica uma diferenciação entre os empregos chamados aqui de constativos e os empregos metaenunciativos:

- (5) *À partir de ces données, j'observe que X est plus grand que Y.*
[À partir desses dados, eu observo que X é maior que Y.]

Dada essa escolha de partida, restringimos nossa observação a verbos prototipicamente utilizados para a constatação científica, nas construções com o *nous* ou o *on* ("*Nous constatons (que)*", *nous constatons* + SN) que assinalam a constatação feita pelo autor cientista. Como o presente do indicativo, na forma ativa, é privilegiado em larga escala nesses empregos, limitamos nossa investigação a essa forma verbal¹⁷. As variantes modais com *pouvoir* [poder] podem ter diferentes interpretações que terão de ser examinadas à luz dos dados do *corpus*. Com relação às diferenças relacionadas aos índices pessoais (*nous* vs. *on*), partimos da ideia de que o *on* se encontra com mais frequência na coconstatação, e de que, nesse caso, a estrutura constativa endossa ainda uma função dialógica.

¹⁴ N.T.: refere-se à classe gramatical do francês *proposition subordonnée* complétive: uma oração que depende de outra principal e que normalmente exerce o papel de objeto direto. Divide-se em Conjuntiva, Infinitiva e Interrogativa.

¹⁵ Como assinalado por TUTIN (2010a), o autor singular é preferencialmente representado pelo *nous* e mais raramente pelo *on*, permanecendo o *nous* de modéstia como norma nos textos científicos.

¹⁶ Encontramos algumas ocorrências no *corpus* Scientext, como esta: "entre os diferentes tipos de interações comunicativas que podem ser produzidas em situações de aprendizagem, constato que as interações argumentativas e explicativas são particularmente propícias à coelaboração de noções". {psy-hdr-35-body}

¹⁷ Outras formas poderiam ter sido estudadas, sobretudo a forma passiva: *il est constaté que...* [constata-se que...]; formas de valor injuntivo como *il est à noter que* [deve-se observar que], *il faut noter que* [é preciso destacar que] ou o imperativo de primeira pessoa: *constatons que...* [constatemos que...], que podem ser relevantes em certos contextos, mas que não puderam ser abordadas no presente estudo.

2.2. Os verbos estudados

Para selecionar os sinônimos, partimos do Dicionário Online CRISCO¹⁸ (*Centre de Recherches Inter-langues sur la Signification en Contexte*), composto com base em dicionários clássicos e não em textos científicos. Esse dicionário apresenta como primeiros sinônimos: *voir* [ver]; *remarquer* [observar], *apercevoir* [perceber]; *reconnaître* [reconhecer], *découvrir* [descobrir], *éprouver* [experimental]; *connaître* [conhecer], *vérifier* [verificar]; *observer* [observar], *s'apercevoir* [aperceber-se]. De pronto, eliminamos alguns desses verbos, como *connaître*, *reconnaître* e *éprouver*, por serem excessivamente polissêmicos e, por vezes, terem sentidos muito distantes da constatação científica. Mantivemos *s'apercevoir*, que apresenta algumas especificidades interessantes, mas eliminamos *apercevoir* por ser muito pouco frequente na escrita científica. Por outro lado, adicionamos dois verbos que aparecem na listagem em um lugar de menor destaque, mas que nos parecem interessantes para a escrita científica, na medida em que implicam uma intencionalidade mais forte: o verbo *noter* [observar], que tem uma motivação semântica diferente, mas um emprego comparável, e o verbo *remarquer* [observar], muito próximo, devido ao uso e à etimologia, dos verbos de percepção (no século XVI, ainda significava, segundo o TLFi: "deter seu olhar, sua atenção em algo"). Não incluímos, entre nossos constativos, verbos "de prova" que pressupõem uma constatação: assim como *vérifier*, há *prouver* [provar], *démontrer* [demonstrar] ou ainda *établir* [estabelecer], que, além de tudo, têm a inconveniência de serem polissêmicos. Esses verbos, sem dúvida relacionados estreitamente à nossa problemática, merecem um estudo específico. Selecionamos os empregos correspondentes ao sentido de constatação, eliminando manualmente¹⁹ os enunciados nos quais esses verbos assumem um sentido diferente. Em síntese, para cada verbo, consideramos: a forma constativa "simples", a forma modalizada com *pouvoir*; a forma com *nous*, a forma com *on*, e examinamos as diferentes construções. Para o verbo *voir*, foi feito um exame mais aprofundado dessas construções; e para os demais, nos quais pudemos observar previamente que, no emprego constativo, se dividem essencialmente entre completiva / SN, nós nos debruçaremos unicamente sobre essa divisão. A Figura 1 apresenta um primeiro levantamento, limitado ao número de ocorrências encontradas para cada um dos verbos estudados (após desambiguação e seleção do sentido constativo): *observer* e *constater* se destacam com mais de 400 ocorrências, seguidos de *voir*, *remarquer* e *noter* (mais de 200 ocorrências); e, por fim, *s'apercevoir*, do qual encontramos apenas 23 ocorrências relevantes.

¹⁸ N.T.: grupo de pesquisa especializado na análise da articulação entre sintaxe e semântica, criado em 2000 na Universidade de Caen-Normandia, França.

¹⁹ Efetuamos as contagens a partir de diferentes construções pertinentes, referindo-nos à lista de lemas para garantir que não esquecemos nada.

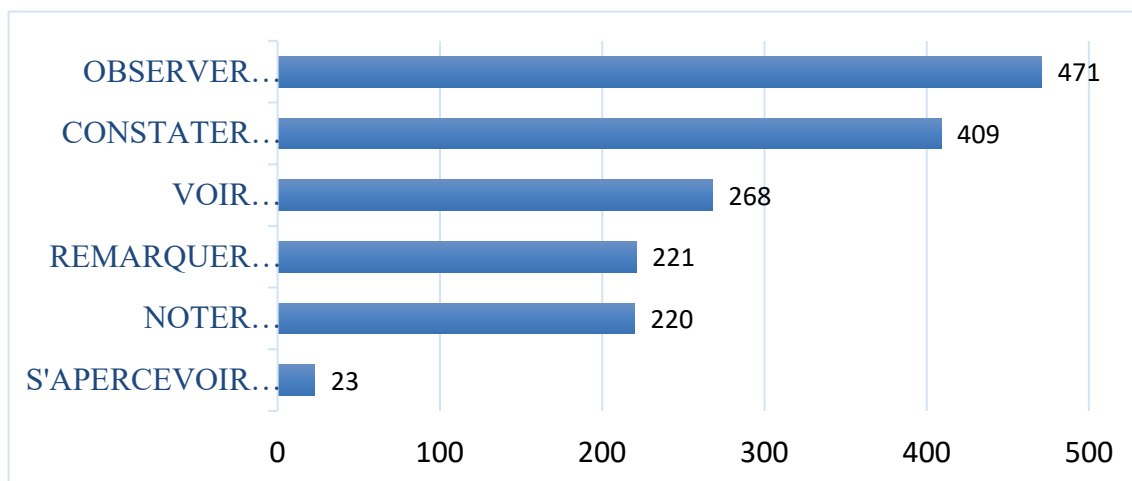


Figura 1: Número de ocorrências para cada um dos verbos estudados

2.3. Estruturas sintáticas e empregos dos verbos de constatação

Em um estudo anterior (Grossmann e Tutin, 2010b), havíamos evidenciado algumas características de *voir* com inciso (*on l'a vu*) [viu-se] ou em estruturas com *comme* (*comme on l'a vu...*) [como se viu]. Percebemos que essas construções, com um valor metaenunciativo que obriga o leitor a concordar, praticamente não eram mobilizadas com outros verbos (*observer*, *constater*, *remarquer*). O verbo *observer*, por sua vez, pediu um complemento nominal com mais frequência que os outros. Pelas razões explicadas acima, deixaremos de lado o inciso (que apresenta forma constativa "fraca"), assim como as estruturas que utilizam *comme*, que têm características específicas, e nos centraremos nos enunciados constativos "fortes". Para os verbos estudados, a estrutura privilegiada dessa constatação no sentido forte é uma construção com uma completiva conjuntiva (*nous voyons que la température remonte*) [nós vemos que a temperatura aumenta] ou com complemento nominal na maioria das vezes deverbal (*nous constatons une légère remontée de la température*) [nós constatamos um ligeiro aumento da temperatura]. São essas duas construções que observamos prioritariamente para cada um dos verbos, sem deixar de fora outras, particularmente com *voir*, o mais bem provido nesse aspecto.

2.4. O verbo *voir* como verbo de constatação

O verbo *voir* (utilizado no presente com *on* e *nous*, com e sem o verbo modal *pouvoir*) se caracteriza por uma frequência bastante alta (está na terceira posição), mas sobretudo pela diversidade de seus empregos e de suas construções. No *corpus* do Scientext, o emprego com o indefinido *on* domina amplamente, com o modal permanecendo muito pouco frequente:

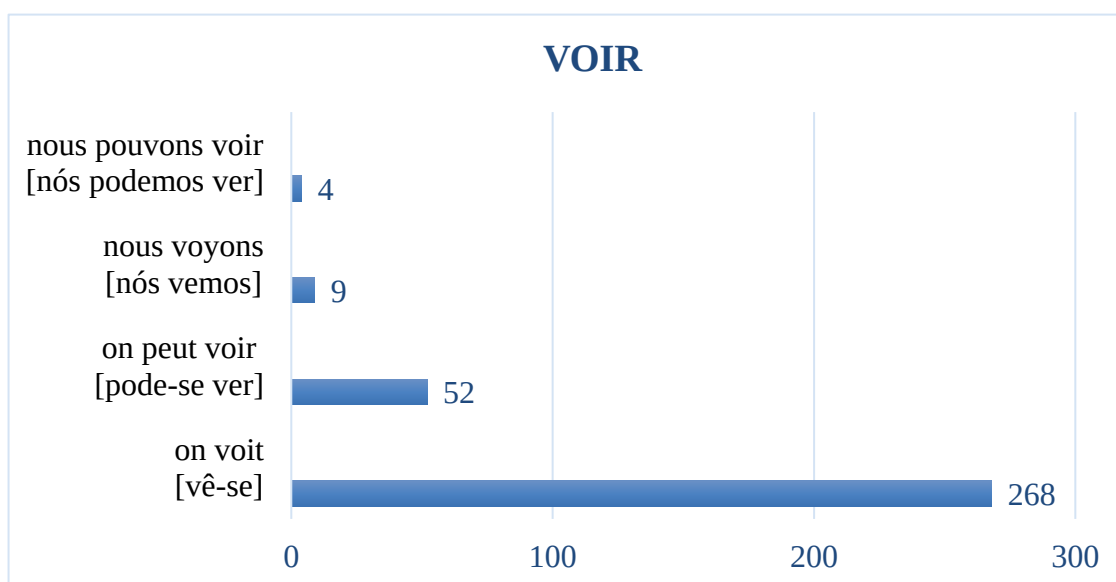


Figura 2: Distribuição para VOIR (n = 333)

Além de estruturas com *comme* ou sem o *comme* com inciso do tipo *on le voit* [vê-se], que são deixadas de fora aqui²⁰, na maioria das vezes esse verbo introduz uma oração completiva conjuntiva, forma prototípica de assinalar uma constatação que queremos compartilhar com o leitor.

Encontramos também com o verbo *voir*:

- interrogativas indiretas, usadas sobretudo para marcar o grau elevado (*on voit à quel point, on voit combien...*)
[vê-se a qual ponto, vê-se o quanto];
- orações infinitivas, muitas vezes recorrendo a verbos com valor aspectual incoativo, do tipo *apparaître, émerger, poindre, se désigner, surgir, se profiler* [aparecer, emergir, despontar, desenhar-se, surgir, configurar-se]:

(6) *Mais parallèlement, nous voyons une seconde tendance émerger entre N3 et N5, que nous appellerons morphographique.* {lin-hdr-91-body}
[Mas, paralelamente, nós vemos uma segunda tendência emergir entre N3 e N5, que chamaremos de morfográfica.]

Algumas delas podem ter um valor avaliativo (em particular com *on le voit bien*), em função do sintagma nominal, sujeito da oração infinitiva:

²⁰ Esse tipo de emprego, já estudado em Grossmann e Tutin (2010a), será aprofundado em uma publicação futura em *Arena Romanistica*. [N.T.: periódico internacional de estudos literários e linguísticos]

(7) *On voit bien ressortir ici les présupposés évidents à la conception...* {sed-the-36-body}

[Pode-se ver sobressaírem claramente aqui os pressupostos óbvios para a concepção...]

A construção infinitiva, depois de um verbo de percepção, é frequentemente descrita como uma forma que favorece uma interpretação menos cognitiva do que a construção completiva. Assim, em (6) o sentido perceptual decerto não fica evidente, mas a construção infinitiva tende a valorizar o próprio ato da constatação; em (7), com o advérbio *bien*, encontra-se um emprego próximo daquele que é analisado logo abaixo, a respeito dos SN com valor avaliativo e, portanto, uma interpretação cognitiva. As construções clivadas no começo das frases desempenham o papel de "ponte" anafórica ou catafórica, introduzindo um exemplo (ou outro dado) que permite ilustrar ou confirmar o que acaba de ser dito:

(8) *C'est ce qu'on voit dans cet extrait...* {lin-the-47-body}

[É o que se vê neste trecho...]

Por fim, encontra-se um número significativo de SN na função de objeto. Nesse tipo de emprego, o SN é geralmente abstrato e pode ter um valor avaliativo de acordo com os complementos nominais utilizados:

(9) *On voit rapidement les limites des deux méthodes présentées précédemment.* {ele-the-110-body}

[Veem-se rapidamente os limites dos dois métodos apresentados anteriormente.]

O que "se vê" nem sempre é um resultado, mas, às vezes, um elemento que permite uma consideração sobre a pesquisa em si, tendo, dessa forma, um valor metatextual. Assim, podem-se encontrar outros tipos de objetos, por exemplo, quando se trata de sintetizar resultados: no exemplo (10), o uso conjunto de *observer* demonstra bem a complementaridade dos dois verbos:

(10) *Toutefois, si on observe les adjonctions de marques plurielles à l'infinitif, on voit trois tendances, valides à tous les niveaux scolaires...* {lin-hdr-599-body}

[Entretanto, caso se observem as adições de marcas plurais no infinitivo, vemos três tendências, válidas para todos os níveis escolares...]

Com o clítico anafórico *y*, a estrutura com *voir* permite apresentar uma interpretação proveniente dos dados anteriormente apresentados:

(11) *On peut y voir l'effet de certaines pratiques idiomatices dans la mesure où...* {lin-the-27-body}

[Pode-se ver nisso o efeito de certas práticas idiomáticas na medida em que...]

Da mesma forma, a estrutura com o clítico *en* é utilizada para adicionar um exemplo ilustrando o que acaba de ser dito:

(12) *On en voit en particulier un exemple avec les circonstancielles "si... alors" à valeur générique [...]* {lin-the-27-body}

[Vê-se disso em particular um exemplo com as adverbiais *se ... então* com valor genérico [...]]

Por último, se nota também, mais raramente, o uso de participípios passados com valor passivo:

(13) *On voit donc clairement illustré ici le principe d'équivalence fonctionnelle.*{lin-hdr-604-body}

[Vê-se então claramente ilustrado aqui o princípio da equivalência funcional.]

A figura a seguir mostra a distribuição das principais construções constativas para *voir*²¹.

²¹ Estão incluídas as estruturas que compreendem o verbo modal *pouvoir*: *nous pouvons voir* [nós podemos ver], *on peut voir* [pode-se ver], assim como os empregos com os clíticos *le* ou *y* (*nous pouvons le voir* [nós podemos vê-lo], *nous pouvons y voir* [nós podemos ver isso]), quando eles não fazem parte de uma estrutura com inciso. O verbo modal *devoir* [dever] tampouco foi levado em conta porque ele nos parece escapar, em grande medida, da interpretação constativa.

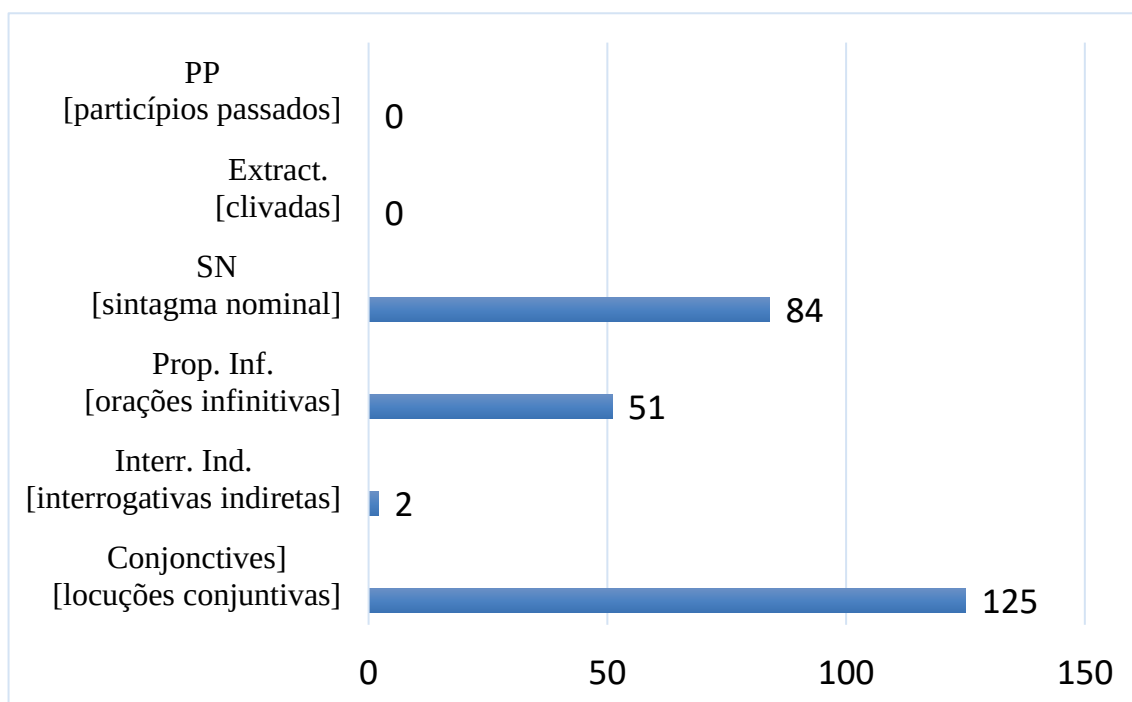


Figura 3: Distribuição das construções de *voir* (após *nous voyons* e *on voit*) (n = 260)

As construções com orações infinitivas e com participios passados com valor passivo são específicas do verbo *voir*, não aparecendo com os demais verbos. Evidentemente, este não é o caso das locuções conjuntivas, muito numerosas, seja qual for o verbo de constatação. Mas para quais fins as conjuntivas depois de *voir* são utilizadas? Dois tipos de empregos se destacam no *corpus*:

- a constatação conclusiva (após um desenvolvimento argumentativo), que propõe uma interpretação dos resultados obtidos:

(14) *On voit donc que, tout bien pesé, l'orthographe du français ne représente pas un cas très satisfaisant d'optimalité sémiographique.* {lin-art-99-body}

[Assim, vê-se que, tudo ponderado, a ortografia do francês não representa um caso muito satisfatório de otimização semiográfica.]

A constatação conclusiva, resultante do desenvolvimento argumentativo que a antecede, não precisa necessariamente apoiar-se em uma posição precisa no texto. No entanto, é bastante frequente que um adjunto adverbial de lugar seja utilizado (*on voit ici... on voit par là que...*). Esse adjunto com valor anafórico parece aproximar esse emprego do funcionamento das estruturas com *comme*. Porém, trata-se de uma constatação no sentido forte do termo, uma vez que é feita a partir dos elementos que o antecedem, e não

simplesmente evocada, como no funcionamento de *voir* como inciso. A utilização de *on*, claramente privilegiada na constatação conclusiva, parece marcar a vontade de incluir o leitor.

- a constatação de observação compartilhada ou coconstatação (Grossmann e Tutin, 2010a, 2010b): baseia-se em um chamado à observação de uma figura, um gráfico ou um exemplo ou quaisquer outros dados como suporte para a demonstração (geralmente indicados por um dêitico (*ici, là...*) [aqui, lá...] ou por um sintagma preposicional), convidando o leitor a consultá-los:

(15) *On voit clairement sur ce graphique que la seule présence d'une bouche souriante quelle que soit l'expression dans laquelle cette bouche est insérée, correspond à une augmentation nette du pourcentage de reconnaissance.*
{psy-com-112-body}

[Fica claro neste gráfico que a simples presença de uma boca sorridente, independentemente da expressão em que essa boca está inserida, corresponde a um claro aumento na porcentagem de reconhecimento.]

Lembremos que, no caso da constatação de observação compartilhada, frequentemente se encontra um localizador que permite remissão a um elemento peritextual (figura, gráfico) ou a uma porção de texto semiograficamente identificável (exemplo numerado, parágrafo ou parte do texto, etc.). Isso corresponde à estrutura evidenciada no trabalho anterior (Grossmann & Tutin, 2010a) que identifica três elementos de um mesmo "quadro" (ou *frame*, na acepção do Fillmore et al., 2003): a testemunha, linguisticamente expressa através dos pronomes inclusivos *nous* e *on*; o fato, objeto da constatação, expresso pelos vários meios já mencionados, mas especialmente pela locução conjuntiva e pelo SN objeto; e a localização, às vezes implícita (o autor cientista refere-se ao seu estudo ou pesquisa), mas que também pode remeter explicitamente, como acabamos de ver, a um elemento peritextual (gráfico, esquema, nota), a um exemplo ou a uma seção do texto. Para além desses dois empregos principais, *voir* também permite fazer avaliações ou comentários, ou mesmo trazer esclarecimentos como este emprego com *ici*, por exemplo, fornecido em nota:

(16) *On voit ici que le modèle de la mémoire demanderait à être précisé : les représentations d'individus ou catégories évoquées par les symboles ne devraient pas avoir le même statut que celles effectivement perçues.* {tal-the-84-notes}

[Vê-se aqui que o modelo da memória exigiria ser definido: as representações de indivíduos ou categorias evocadas por símbolos não devem ter o mesmo *status* daqueles realmente percebidos.]

Trata-se aqui de antecipar as possíveis objeções do leitor, como parte da estratégia argumentativa, uma vez que o *on voit que* marca também a convivência com ele, pressupondo sua adesão ao que é dito. A plasticidade de *voir*, a própria diversidade de construções em que ele cabe o deixam longe de ser o melhor representante da classe de constativos no sentido estrito do termo: é uma ferramenta que serve para (quase) tudo: coconstatação, constatação conclusiva, marcador avaliativo (*on voit bien...*), marcador de inferência, marcador de pressuposição²².

2.5. O verbo *observer*

Observer aparece como o mais frequente dos verbos estudados (com 471 ocorrências relevantes, incluindo as formas com *on*, *nous* e a modalização com *POUVOIR*). O *on* indefinido é novamente privilegiado, porém de maneira menos nítida do que com *voir*, com o emprego da primeira pessoa do plural sendo bastante alto (fig.4)

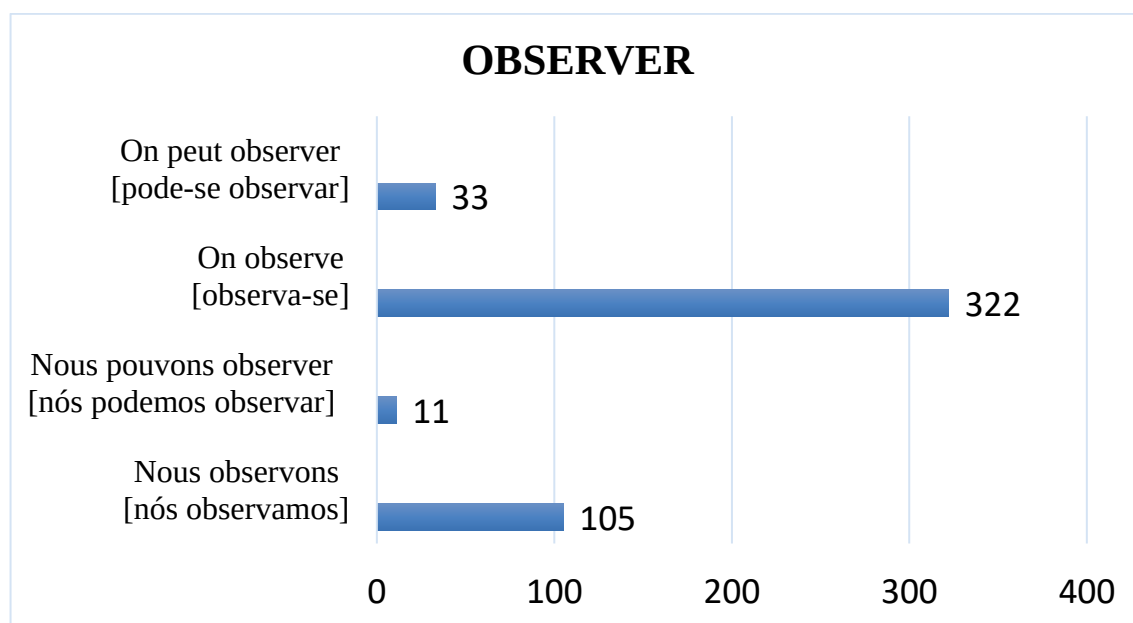


Figura 4: Distribuição dos empregos indefinidos e pessoais, modais e assertivos (n = 471)

Observer tem um uso com *nous* relativamente mais frequente que *voir*²³, permanecendo o emprego modal com *pouvoir* bastante raro. Em comparação com outros verbos de constatação, *observer* tem uma característica que se destaca: é o único a

²² Note-se que há um emprego específico de *voir* como um verbo de julgamento, impossível com os outros verbos aqui estudados (encontrado apenas também com *considérer*), na estrutura com atributo essencial de objeto: *On peut voir ces acteurs comme une matérialisation de connexions réentrantes qui...* [tal-the-5-notes] [Pode-se ver esses atores como uma materialização de conexões reentrantes que...] Trata-se aqui de um efeito de qualificação ou de aproximação (FUCHS e GOFFIC, 2005) distante do sentido constativo que nos interessa.

²³ O emprego do *nous* mereceria aqui um estudo mais detalhado: além do uso do *nous* acadêmico do autor, frequentemente encontram-se empregos inclusivos com *observer*.

privilegiar – e de maneira muito clara – a construção com um complemento nominal (277 ocorrências de SN contra 134 **Que_O**²⁴). Os objetos nominais podem ser nomes de processos (diluição, aumento) ou assinalar um resultado (tendências, disparidades). Dois tipos de actante são privilegiados, o primeiro introduzindo um SN que marca o resultado de um processo:

- (17) *Nous observons **une** déstabilisation **énergétique** de 25 Kcal/mol...* {ele-the-651-body}
[Observamos uma desestabilização energética de 25 Kcal / mol...]

Observer é seguido por um nome predicativo marcando o resultado de um processo, podendo o SN ser parafraseado por uma estrutura do tipo "*qu'il se produit*": *Nous observons qu'il se produit une déstabilisation énergétique...* [Nós observamos que se produz uma desestabilização energética...]

Em (18), o mesmo tipo de informação é obtido graças a uma estrutura **Que_O** cujo sujeito gramatical é um nome de processo:

- (18) *Nous observons que **la variance** (S2) **diminue significativement** de T1 à T2...* {lin-com-111-body}
[Nós observamos que a variação (S2) diminui significativamente de T1 para T2...]

As estruturas **Que_O** favorecem as constatações do tipo generalizante:

- (19) *En ce qui concerne le clip Hotdog, on observe **que tous les énoncés contiennent des adverbiaux temporels**...* {lin-com-69-body}
[Com relação ao clipe Hotdog, observa-se que todos os enunciados contêm adjuntos adverbiais temporais...]

A estrutura completiva permite tanto a constatação relacionada à descoberta do processo (o que também permite a estrutura com SN) quanto a constatação generalizante. No entanto, ela é relativamente menos representada do que para outros verbos de constatação. Em compensação, o verbo *observer* é bastante utilizado com um locativo, construção em que a motivação visual retoma sua eficácia, como mostram os exemplos 20 e 21:

- (20) *On observe **sur la Figure 58** que les activités de réparation...* {bio-the-95-body}
[Observa-se na figura 58 que as atividades de reparo...]

²⁴ N.T.: estrutura com "que" que introduz uma oração subsequente (nomenclatura do autor).

(21) *Dans un noyau inter-phrasique, on peut observer par microscopie optique deux types de structures chromatiniennes...* {bio-the-281-introduction}

[Em um núcleo interfásico, pode-se observar por microscopia óptica dois tipos de estruturas de cromatina...]

No exemplo (20), tem-se um caso claro de coconstatação, que envolve o leitor no exame de uma figura. Em (21), o sentido constativo se atenua em favor de um direcionamento para a observação "instrumental" do microscópio.

2.6. O verbo *s'apercevoir*

Este verbo tem uma baixa frequência no *corpus* (23 ocorrências, 20 delas com o indefinido: *on s'aperçoit* [percebe-se]). Não há registro da forma modalizada. Já o quase sinônimo *se rendre compte* [dar-se conta], de emprego muito próximo, apresenta uma dezena de ocorrências no *corpus*. A estrutura utilizada é quase exclusivamente com *Que_O* (há uma única ocorrência com *SN*). O verbo *s'apercevoir* é utilizado para assinalar que um fato se revelou ao pesquisador no final do processo de observação. Ele é, portanto, sempre perfectivo, ao contrário de *voir*, que pode ser perfectivo ou imperfectivo. Isso explica por que, como já observamos na relação *voir/observer*, o uso de *s'apercevoir* é muitas vezes feito concomitantemente, no contexto que o antecede, com um verbo de observação como *considérer* (em 22) ou *observer* (em 23):

(22) *Si l'on considère des acides aminés individuels, seuls ou dans de très petits peptides, on s'aperçoit que chacun peut occuper de l'ordre de dix conformations...* {bio-art-184-introduction}

[Considerando-se os aminoácidos individuais, isoladamente ou em peptídeos muito pequenos, percebe-se que cada um pode ocupar cerca de dez conformações...]

(23) *En observant de plus près ces productions, on s'aperçoit qu'une consonne prédomine: la consonne /n/...* {lin-the-23-body}

[Observando mais de perto essas produções, percebe-se que uma consoante predomina: a consoante / n /...]

O verbo *s'apercevoir* não funciona bem com um adjunto adverbial de lugar:

(24)? *On s'aperçoit sur la Figure 58 que les activités de réparation...* {bio-the-95-body, modificado}

[? Percebe-se na Figura 58 que as atividades de reparo...]

Por outro lado, como mostrado nos exemplos (22) e (23), ele se adapta perfeitamente a estruturas hipotéticas, ou ao gerúndio, o que torna possível assinalar o processo ao fim do qual a constatação se faz necessária.

2.7. O verbo *noter*

Com 220 ocorrências relevantes encontradas no *corpus*, esse verbo tem uma frequência significativa. A forma com o *on* indefinido é outra vez de longe a mais encontrada:

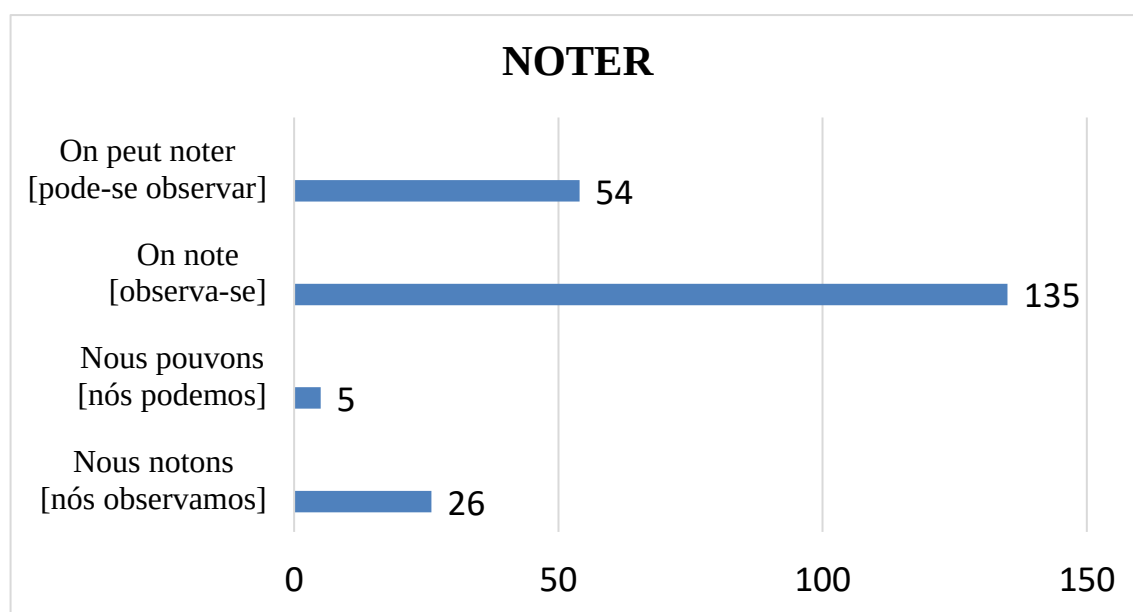


Figura 5: Distribuição de empregos de *noter* (n = 220)

Esse verbo é utilizado em especial para fornecer uma informação interessante embora não central, ou então um elemento adicional que corrobora uma hipótese²⁵.
(em combinação frequente nesses casos com *de plus*, *ici aussi* ...)

(25) *De plus, on note la réapparition de la syllabe qui était ignorée...* {lin-the-23-body}

[Além disso, observa-se a retomada da sílaba ignorada...]

(26) *Ici aussi, on note une alternance entre l'interrogation des élèves à forte...* {sed-com-217-body}

[Aqui também, observa-se uma alternância entre a interrogação dos alunos com forte...]

²⁵ Eliminamos todos os empregos gráficos (exemplo: *noter un message* [no sentido de registrar]) que não têm sentido constativo.

Isso explica por que esse verbo é frequentemente utilizado com o modalizador *pouvoir* (59 de 220 ocorrências), *on peut noter* introduzindo um comentário ou propiciando destacar de um todo algum elemento importante ou característico:

(27) *Parmi les projets de grande envergure, on peut noter WordNet...* {lin-com-66-introduction}

[Entre os projetos de grande escala, pode-se destacar o WordNet...]

O verbo *noter*, assim como *observer* e *voir*, pode ser utilizado para a coconstatação, por exemplo, quando se trata de ressaltar uma tendência visível em um gráfico:

(28) *Sur cette représentation, on note les différentes atténuations et retards des trajets multiples.* {ele-the-9-body}

[Nesta representação, destacam-se as diferentes atenuações e atrasos de percursos múltiplos.]

No entanto, esse uso com locativo permanece periférico, encontramos poucos advérbios de lugar nos enunciados com *noter*. Enfim, mesmo predominando a construção completiva, as construções nominais são bem representadas, geralmente com substantivos predicativos traduzindo um processo.

2.8. O verbo *constater*

O verbo *constater* é frequente no texto científico (409 ocorrências em nosso *corpus*). O emprego modal com *pouvoir* é fraco (apenas 26 ocorrências). O emprego pessoal com o *nous* acadêmico está relativamente presente (aproximadamente um quarto dos empregos encontrados).

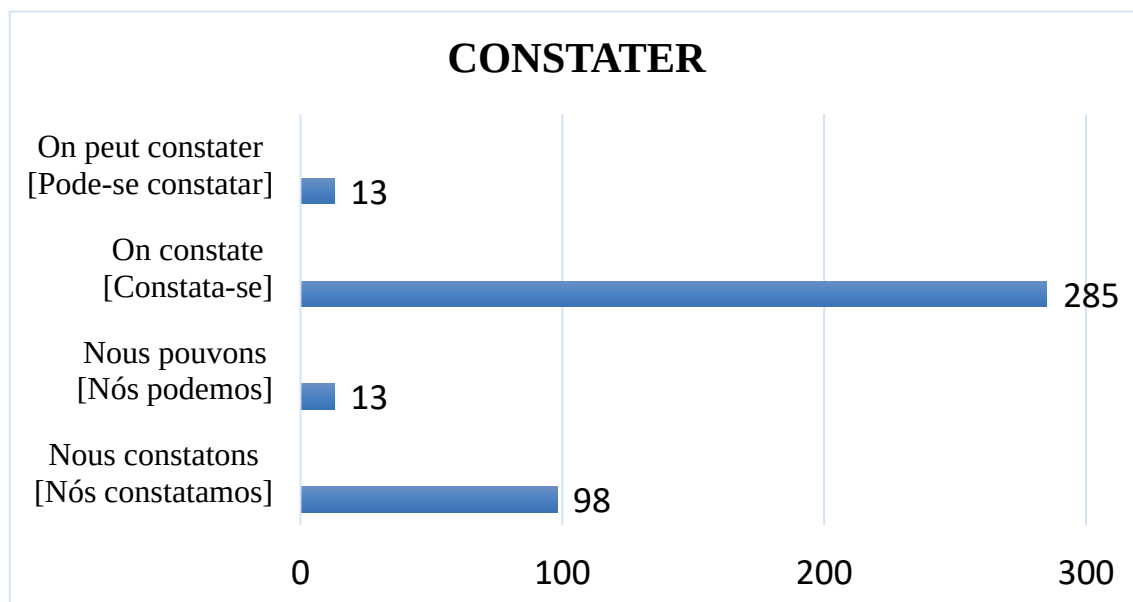


Figura 6: Distribuição de empregos de *constater* (n = 409)

No que diz respeito às estruturas sintáticas privilegiadas, a estrutura com *Que_O* chega novamente disparada na frente (314 ocorrências contra 77 da construção com SN). Semanticamente, os complementos nominais permitidos pelo verbo *constater* são um pouco diferentes dos complementos de *observer*, *noter* ou *voir*. Embora seja possível constatar a existência ou a inexistência de um fenômeno, parece mais difícil ver essa existência ou observá-la²⁶ (pode-se a rigor destacá-la, mesmo que nós não tenhamos encontrado ocorrências desse tipo no *corpus*). O verbo *constater* tem uma dimensão mais abstrata. É o que também explica que ele dificilmente possa ser substituído por *observer* quando este último tem um emprego empírico (e com um nome que não seja o do processo). Compare (29) e (30):

(29) *les seuls indices de structuration sont les paragraphes, on observe des textes avec des paragraphes extrêmement longs...*{lin-the-27-body}

[os únicos indícios de estruturação são os parágrafos, observam-se textos com parágrafos extremamente longos...]

(30) **les seuls indices de structuration sont des paragraphes, on constate des textes des paragraphes extrêmement longs...*{lin-the-27-body, modificado}

[*os únicos indícios de estruturação são parágrafos, constatam-se textos de parágrafos extremamente longos...]

²⁶ Encontramos, em compensação: "On observe la présence de structures tourbillonnaires dans la zone... {mec-the-108-body}" [Observa-se a presença de estruturas de vórtices na área...]. Entretanto, *présence* é menos abstrato do que *existence*.

2.9. O verbo *remarquer*

O verbo *remarquer* tem uma frequência média, com uma representação bastante forte dos empregos que mobilizam o *nous* acadêmico (pouco mais de um terço).

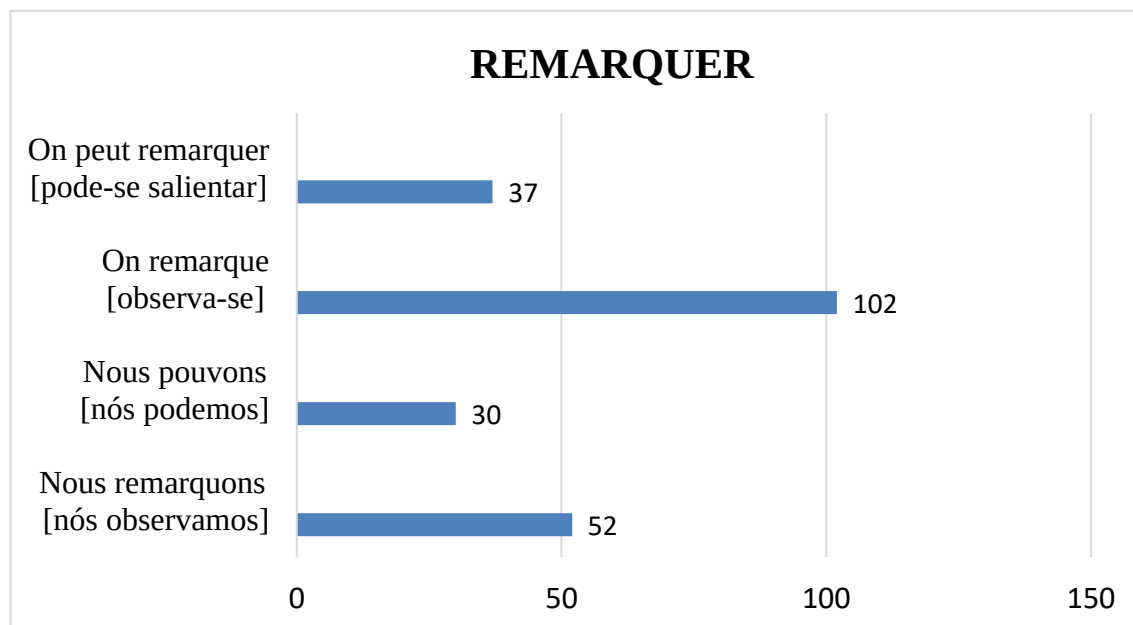


Figura 7: Distribuição de empregos indefinidos e pessoais, modais e assertivos de *remarquer* (n = 221)

Encontra-se uma frequência razoavelmente alta do emprego com o modalizador *pouvoir* (67 ocorrências), o que se explica pelo fato de o verbo *remarquer* permitir a introdução de um comentário. Desse ponto de vista, ele se aproxima de *noter* que, como ele, coocorre frequentemente com *également* ou *aussi*:

- (31) *On peut remarquer également que l'accès à des pratiques personnalisées...*{sed-com-171-conclusion}
[Pode-se observar igualmente que o acesso a práticas personalizadas...]

No entanto, como vimos, ao contrário de *noter*, o verbo *remarquer* aparece com bastante frequência com o *nous* acadêmico, favorecendo um emprego mais monológico do que dialógico. Finalmente, do ponto de vista sintático, a construção com *Que_O* é mais uma vez claramente privilegiada (197 empregos contra 23 com *SN* e 1 com outra). Os complementos circunstanciais de lugar, embora não tão frequentes como com *observer*, estão presentes.

3. Conclusão

Partimos da hipótese de que havia uma classe de verbos "constativos" que permitia, nos gêneros científicos, formular uma afirmação científica, em geral sustentada por

elementos aos quais o leitor do texto possa se referir. A identificação dos enunciados constativos apresenta um certo interesse na medida, especialmente, que tornam consistentes os resultados obtidos ou representam marcos importantes do raciocínio científico.

De maneira geral, encontram-se muitos pontos em comum entre eles, mas também diferenças de funcionamento e de uso. Entre as convergências, há o fato de que todos os verbos estudados – com exceção de *s'apercevoir*, que utiliza quase exclusivamente a construção com Que_O – constroem seu objeto permitindo a alternância entre as construções Que_O e SN objeto. A frequência da estrutura com Que_O, e o papel da modalização epistêmica que eles desempenham no enunciado os aproximam dos verbos por vezes chamados indevidamente de "assertivos"²⁷.

Entre os verbos estudados, dois se destacam claramente, tanto por sua frequência como por sua centralidade: *observer* e *constater*, sendo o primeiro, em razão de seu semantismo, mobilizado mais vezes para descrever a fase empírica da pesquisa, enquanto o segundo, mais abstrato, para traduzir os resultados. Em muitos contextos, no entanto, os dois verbos são intercambiáveis, *observer* podendo assumir os valores imperfectivos ou perfectivos assim como *constater*. Este último é mais frequentemente utilizado com o *nous* acadêmico do que o anterior; ele também é mais empregado no que chamamos de constatação conclusiva. O *nous* utilizado parece raramente inclusivo: trata-se do *nous* do autor que tira suas conclusões. Os empregos modalizados com *pouvoir* são raros. *Observer*, por sua vez, aparece mais vinculado à coconstatação (podendo o *nous* ser inclusivo, quando utilizado).

Noter e *remarquer* parecem ser semanticamente bem próximos, pois estabelecem uma constatação incidental, tendo ocasionalmente valor de acréscimo, de comentário ou de ilustração para uma afirmação anterior. Ambos são frequentemente associados a advérbios do tipo *d'ailleurs*, *pourtant*, *néanmoins*, *aussi* [ademais, no entanto, não obstante, também], o que parece indicar que eles têm fundamentalmente um papel argumentativo e marcam por vezes um contra-argumento e, por outras, um acréscimo que reforça o argumento anterior. Dado o seu emprego mais frequente na primeira pessoa, *remarquer* se aproxima de um emprego mais monológico do que *noter*, que tem mais afinidades com a coconstatação implicando o leitor, embora aqui se trate novamente de simples tendências, já que os dois empregos são sempre possíveis.

Voir e *s'apercevoir* têm em comum o fato de apontarem mais para o aspecto interpretativo da constatação. *Voir*, no entanto, tem uma gama muito ampla de uso: alguns de seus empregos conduzindo-o a sair do papel de constativo para adquirir um valor avaliativo (cf. *on voit bien*), o que exigiria certamente uma triagem mais acurada do que a que realizamos; *s'apercevoir*, em compensação, é frequentemente utilizado apenas para marcar a conclusão de um processo de observação.

²⁷ Veja a este respeito as considerações de Borillo (1982).

Para todos os verbos – com exceção de *constater*, para o qual o número de empregos com *pouvoir* é equivalente –, confirma-se o fato de que a modalização com *pouvoir* é mais facilmente utilizada com *on* do que com *nous*, o que parece mostrar suas afinidades com a coconstatação.

Voltamos, para concluir, à noção de constatação assertiva. Equiparamos os comentários metaenunciativos com constativos a uma forma "fraca" de afirmação: quando o verbo de constatação introduz um sintagma do tipo inciso²⁸, quer em primeira posição, posição intermediária ou posterior. Borillo (1982, p.35) assinala que, em posição posterior, não é possível refutar um enunciado como (33):

- (32) Tu est en retard, je constate.
[Você está atrasado, eu constato].
(33) * Tu es en retard, je ne constate pas.
[*Você está atrasado, eu não constato.]

É verdade também que é igualmente difícil, nos gêneros científicos, refutar uma estrutura com *comme*, como mostram os exemplos inventados (34) e (35):

- (34) ? X est similaire à Y, comme nous ne l'avons pas vu précédemment. [?
X é semelhante a Y, como nós não o vimos antes.]
(35) ? X est similaire à Y, comme nous ne le voyons pas sur la figure Z. [X é
semelhante a Y, como nós não o vemos na Figura Z.]

Constata-se, no entanto, a proximidade funcional desses enunciados com a forma que compreende a constatação forte:

- (36) Nous voyons sur la figure Z que X est similaire à Y
[Nós vemos na Figura Z que X é semelhante a Y.]

Nesses exemplos, as restrições parecem surgir mais da pragmática textual do que do nível enunciativo: com efeito, a negação de (36) seria tão estranha ao leitor quanto os exemplos apresentados por Borillo. Sem dúvida, devemos continuar explorando mais as formas de coconstatação, reservando uma atenção especial à coconstatação no presente, com locativo expresso: as estruturas com *comme*, neste caso, funcionam bem, ao que tudo indica, como constatações assertivas. Mesmo que seja possível manter a distinção entre incisos "retóricos" ou com função de marcação textual – cujo verbo está geralmente no passado composto – e os enunciados plenamente constativos, percebe-se que a fronteira

²⁸ Não entraremos aqui no debate terminológico sobre os incisos, incidentes e outras estruturas metaenunciativas. Pensamos, como Wilmet (1998, p.574), que o critério enunciativo é um elemento importante para definir o inciso, o que funcionalmente aproxima as estruturas com *comme* dos incisos.

que separa a constatação assertiva da constatação "fraca" é por vezes tênue. Por um lado, alguns verbos (*noter, remarquer*) traduzem seguidamente uma constatação incidental. Por outro, certas estruturas coconstativas com *comme*, com o verbo no presente, contribuem, do mesmo modo que a forma completiva ou o complemento nominal, para estabelecer a constatação científica.

Referências

- AUTHIER-REVUZ, J. (1995). *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris, Larousse.
- BORILLO, A. (1982) Deux aspects de la modalité assertive : croire et savoir, *Langages*, 67, 33-53.
- CAPPEAU, P.; SAVELLI, M. (2001) C'est bien comme ça ? Étude des constructions en *comme*, *Recherches sur le français parlé*, 16, 39-62.
- DEBAISIEUX, J.-M.; MARTIN, PH. (2010) *Les parenthèses, étude macrosyntaxique et prosodique sur corpus*, in *La parataxe, structures, marquages et exploitations discursives*, Berne, pp. 307-338, Peter Lang.
- DELPLANQUE, A. (2006) Juger d'après les apparences : le cas du français, *Corela*, Les verbes d'apparence. Publié en ligne le 13 mars 2006. URL : <http://corela.edel.univ-poitiers.fr/index.php?id=1284>. Consulté le 31/05/2012.
- FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. (2003) Background to Framenet. *International Journal of Lexicography*, 16(3), 235–250.
- FUCHS, C.; LE GOFFIC, P. (2005) La polysémie de *comme*. In Olivier Soutet (dir.), *La polysémie*, pp. 267-291, Paris, Presses de l'Université Paris Sorbonne.
- GROSSMANN, F.; TUTIN, A., (2010), Evidential Markers in French Scientific Writing: the Case of the French Verb *voir*. In Smirnova E., Diewald, G. (eds). *Evidentiality in European Languages*, Berlin, New York, Mouton de Gruyter.
- GROSSMANN, F.; TUTIN, A. (2010). Les marqueurs verbaux de constat : un lieu de dialogisme dans l'écrit scientifique. Actes du colloque Dialogisme : langue, discours 2010, 8-10 septembre 2010. Montpellier, Praxiling. Disponible à l'adresse : http://www.univ-montp3.fr/praxiling/IMG/pdf_GrossmannTutin.pdf
- WILMET, M. (1998, 2^{ème} éd.), *Grammaire critique du français*, Paris, Bruxelles, Hachette et Duculot.

Como citar este texto (ABNT):

GROSSMANN, F. Tradução de Alex de Cássio da Silva. Os verbos de constatação no texto científico. *Cadernos de Tradução*, Porto Alegre, n. 43, jul/dez, p. 149-169, 2018.

A FRASEOLOGIA TRANSDISCIPLINAR DOS TEXTOS CIENTÍFICOS: DAS COLOCAÇÕES ÀS ROTINAS RETÓRICAS¹

Agnès Tutin²

Tradução: Anna Carolina Teles³

Supervisão e revisão: Patrícia Reuillard⁴

Revisão técnica: Sandra Loguercio⁵

1. Introdução

O texto científico tem a ver, sem dúvida, com um socioleto, oriundo de uma comunidade discursiva que compartilha fins comuns (Cf. SWALES, 1990), objetivos retóricos comparáveis e uma sublinguagem específica, caracterizada por traços textuais, pragmáticos, sintáticos e lexicais evidentes (Cf. KOCOUREK, 1991). O artigo científico, por exemplo, sobretudo das ciências duras, é um tipo de texto majoritariamente pré-fabricado, tanto em seu esquema discursivo quanto em suas *expressões fixas* mobilizadas como outros gêneros muito diferentes, a exemplo das cerimônias de casamento, dos anúncios pessoais ou dos processos (GÜLICH e KRAFFT, 1997).

Pode-se considerar que a língua utilizada nesse gênero textual, por vezes chamada de “língua científica geral” (PECMAN, 2004), atravessa, de certa forma, as disciplinas⁶ (Cf. a *linguagem acadêmica* de Biber, [2006]). Essa língua científica geral é ilustrada particularmente por um léxico transdisciplinar, diferente da terminologia dos domínios científicos definidos, “um léxico compartilhado pela comunidade científica, aplicado na descrição e na apresentação da atividade científica [...] que não se refere aos objetos científicos das áreas de especialidade, mas ao discurso sobre os objetos e os procedimentos científicos” (TUTIN, 2007a, p.6). Esse léxico transdisciplinar está principalmente na fraseologia e foi objeto de inúmeros trabalhos de pesquisa, sobretudo em inglês (Cf.

¹ A presente tradução foi autorizada para ser publicada em português, nos Cadernos de Tradução do IL (UFRGS), pela Editora Presses Universitaires de Rennes (PUR) e pela autora Agnès Tutin. Referência bibliográfica completa do artigo original: TUTIN, Agnès. “La phraséologie transdisciplinaire des écrits scientifiques: des collocations aux routines sémantico-rhétoriques”. In: TUTIN, A.; GROSSMANN, F. (org.) *L'écrit scientifique: du lexique au discours*. Rennes: PUR, p. 27-43, 2013.

² Laboratoire de Linguistique et de Didactique du Français Langue Étrangère et Maternelle (LIDILEM), Université Grenoble 3 – Stendhal.

³ Bacharelada em Letras – Tradução Português/Francês, UFRGS.

⁴ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁵ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

⁶ Embora cada família de disciplinas comporte, sem dúvida, especificidades.

HOWARTH, 1996; GLEDHILL, 2000; PECMAN, 2004⁷. Para o francês, Cf. LOFFLER-LAURIAN, 1995). Essa fraseologia tem um papel fundamental no âmbito desse discurso: ela permite primeiramente ao autor/pesquisador reforçar sua pertença a uma comunidade discursiva com o emprego de uma linguagem codificada (*dans la littérature, nous écartons l'hypothèse* [na literatura, nós descartamos a hipótese⁸]...), e facilita, por seu caráter pré-construído, a escrita.

Por outro lado, a fraseologia recorrente nas diferentes disciplinas é uma porta de entrada privilegiada para o estudo epistemológico dos discursos científicos (cf., por exemplo, certos estudos de caso sobre os verbos no âmbito dos projetos KIAP [FLØTTUM *et al.* 2006a] e Scientext [GROSSMANN *et al.*, 2009; TUTIN, 2010b]). Essa fraseologia de gênero pode, além disso, ser explorada na área das ciências da informação, ora como *corpus* de exclusão nos processos de indexação automática⁹, ora como léxico que permite caracterizar o tipo de texto científico¹⁰. Por fim, no âmbito didático, a especificidade linguística desses textos foi amplamente estudada, sobretudo no inglês, na perspectiva do ensino acadêmico, *Inglês para Fins Acadêmicos* (Cf., por exemplo, os trabalhos de SWALES, [1990], HYLAND, [2000], COXHEAD, [2002]). No francês, o interesse pelo *Francês para Objetivos Universitários*, mais recente, responde a objetivos similares, visando aos estudantes estrangeiros que precisam se aculturar na língua acadêmica francesa (Cf. PARPETTE e MANGIANTE, 2010; CAVALLA, 2010).

Nosso interesse pelo estudo linguístico do texto científico e as necessidades mencionadas levaram-nos a refletir sobre o funcionamento da fraseologia transdisciplinar nesse gênero textual. Voltamo-nos, em um primeiro momento, para as colocações, que constituem um subconjunto particularmente produtivo da fraseologia científica (Cf. TUTIN, 2007b). Desejamos, em um segundo momento, expandir esse estudo para além da faceta da fraseologia científica transdisciplinar de uma maneira funcional. Nessa perspectiva, a tipologia elaborada deve servir tanto aos estudos linguísticos e ao

⁷ Destaquemos que a abordagem de Pecman é contrastiva e diz respeito igualmente ao francês.

⁸ N.T.: Os exemplos apresentados em francês e em inglês neste texto recebem traduções em português entre colchetes para que o leitor compreenda o sentido, mas, ao mesmo tempo, possa igualmente compreender a estrutura em francês e/ou em inglês.

⁹ A indexação automática consiste em extrair os termos chave de um texto especializado. A maioria das técnicas baseia-se em procedimentos estatísticos que extraem o léxico específico do texto com ajuda de medida de especificidade. Ora, o léxico transdisciplinar é um léxico específico do gênero, sendo portanto muito frequente nesse tipo textual, mas não se refere aos conceitos da área. Excluir o léxico transdisciplinar permitirá uma melhor indexação dos textos. Vejamos um exemplo em linguística com a palavra *sujeito* [N.T.: em francês, a palavra remete tanto ao *sujeito gramatical*, quanto ao *objeto* de um estudo.]: *sujeito* é um termo, em um texto de linguística, em contextos como *sujeito de uma frase* ou em associação com unidades lexicais como *frase*, *objeto direto*. De outra parte, em *abordar um 'sujet'* [objeto] ou o *'sujet'* [objeto] *abordado*, ele é, sem dúvida, um não-termo. Desse modo, considerar o ambiente lexical das palavras (colocações e outras expressões polilexicais) permite afinar o processo de indexação. Esse tema é abordado no projeto ANR-ContintTermith, coordenado pelo ATILF e do qual participa o LILIDEM.

¹⁰ O artigo é uma revisão do estado da arte, um artigo principalmente do tipo teórico ou do tipo experimental? O léxico é uma porta de entrada privilegiada para caracterizar o tipo do artigo.

estabelecimento de padrões ou modelos para as aplicações de processamento automático previstas¹¹ quanto às aplicações didáticas que queremos desenvolver a longo prazo.

Após apresentar uma tipologia das sequências polilexicais aplicada ao léxico transdisciplinar do texto científico, ocupamo-nos de um tipo de expressão no limite da fraseologia, as rotinas retóricas, de que propomos padrões inspirados nos *frames* semânticos de Fillmore (FILLMORE *et al.*, 2003). Mostramos então como esses *frames* foram implementados em um sistema operacional de processamento de *corpus* no projeto Scientext.

2. Uma tipologia das sequências polilexicais do texto científico

Nosso objeto de estudo concerne à fraseologia do texto científico, que compreende, em sentido amplo, as sequências polilexicais recorrentes e específicas ao gênero texto científico, sejam fixas ou não, e seja qual for o estatuto sintático dos elementos constituintes da expressão. Essa fraseologia, própria ao texto científico, não é uma terminologia, na medida em que não remete aos conceitos chaves da área (por exemplo, a Linguística ou a Geologia). Ela passa até certo ponto pelas disciplinas, mas conjuntos lexicais próprios às famílias de disciplinas podem, sem dúvida, ser extraídos. Essa fraseologia metacientífica e metadiscursiva essencialmente abstrata descreve os procedimentos científicos, o raciocínio científico, a filiação científica, o posicionamento, o metatexto e a estrutura do discurso. Ela não está necessariamente ausente da língua comum, mas pode assumir acepções específicas nesse gênero ou apenas ser muito frequente.

Esse léxico foi objeto de inúmeros trabalhos em inglês, sobretudo no âmbito do *Inglês para Fins Acadêmicos*, sejam trabalhos sobre o discurso acadêmico e científico em uma perspectiva discursiva (HYLAND, 2005), sejam em uma perspectiva textométrica (BIBER, 2006). De mesmo modo, desenvolveram-se trabalhos sobre as colocações, muito produtivas nesse gênero, sejam colocações lexicais (HOWARTH, 1996), sejam enquanto um fenômeno de definição mais ampla (GLEDHILL, 2000). Quanto ao francês, os trabalhos sobre essa fraseologia são menos numerosos, com exceção dos de Pecman (2004; 2007), que detalharemos adiante, assim como estudos mais pontuais sobre as colocações (TUTIN, 2007b). Destaca-se que a fraseologia usada nos textos científicos parece extremamente diversa e que o protótipo da sequência polilexical não-composicional no plano semântico, fixa no plano sintático, e cujos referentes não são atualizados, mostra-se muito limitante. Não se trata apenas de identificar desordenadamente *pacotes lexicais* à maneira de Biber (2006), mas também de caracterizá-los nos planos discursivo, enunciativo, semântico e sintático.

A tipologia que propomos é inspirada ao mesmo tempo em Burguer (1988), Granger & Paquot (2008) e Mel'čuk (2011). O parâmetro mais expressivo é o papel funcional que as expressões cumprem no discurso científico. A tipologia proposta é

¹¹ No âmbito do projeto Scientext, mas também no do projeto ANR Termith na indexação automática.

provavelmente transponível a outros tipos de discurso, no entanto as expressões extraídas são, sem dúvida, próprias ao gênero estudado. A tipologia elaborada deve atender a várias exigências, muitas vezes difíceis de conciliar:

- Deve ser evidentemente delimitada por critérios claros e parecer funcional para o estudo linguístico, em especial para o estudo do discurso científico.
- Deve ser suficientemente intuitiva, em particular para o tratamento semântico, para um tratamento didático.
- Deve ser satisfatoriamente precisa para permitir o estabelecimento de padrões em um Processamento Automático da Linguagem (TAL em francês), que possibilite uma extração informatizada a partir de *corpora* textuais.

A tipologia proposta conduz a quatro tipos de expressões que serão detalhadas a seguir:

1. Sequências polilexicais de função referencial, colocações ou expressões fixas, (ex.: *faire une hypothèse* [levantar uma hipótese]; *point de vue* [ponto de vista])
2. Sequências polilexicais de função discursiva (ex.: *en d'autres termes* [em outras palavras], *pour conclure* [para concluir]...)
3. Sequências polilexicais de função interpessoal, essencialmente expressões de função modal (ex.: *Il est probable* [é provável], *contre toute attente* [contrariando as expectativas]...)
4. Rotinas retóricas, próprias ao discurso científico (ex.: *comme on peut voir sur la figure* [como se pode ver na figura], *le tableau X* [a tabela X]).

2.1. Sequências polilexicais de função referencial

Essas expressões referem-se a conceitos simples ou complexos no texto científico, evocados pelo locutor, sem função discursiva ou comunicativa específica¹². No âmbito de um modelo funcional, como a gramática sistêmico-funcional de Halliday (1994), a metafunção será aqui sobretudo *ideacional*, ou seja, para definir essas expressões não é necessário explicar seu contexto de produção ou sua função discursiva. Do ponto de vista da estrutura semântica interna, podem ser frasemas completos ou locuções completas (MEL'ČUK, 2011), ou seja, expressões como *prendre en compte* [levar em conta], *tenir compte* [ter em conta], *faire partie* [fazer parte], *point de vue* [ponto de vista], que são não-composicionais no plano semântico e muito restritas no plano sintático. Essa classe de expressões inclui, de mesmo modo, as colocações lexicais (expressões recorrentes, essencialmente binárias, cujos elementos apresentam uma relação sintática), como *faire une hypothèse* [levantar uma hipótese], *réfuter une hypothèse* [refutar uma hipótese], *résultats prometteurs* [resultados promissores] etc. (Cf. GROSSMANN e TUTIN, 2003; TUTIN, 2007b). Para ilustrar, apresentamos, no Quadro 1, um subconjunto de colocações transdisciplinares de tipo V-N extraídas de um *corpus* analisado sintaticamente

¹² Esta categoria é, de certo modo, definida pela exclusão das outras.

(expressões mais frequentes em três disciplinas científicas: medicina, linguística e economia). Nota-se que muitas dessas expressões não são exclusivas dos textos científicos, mesmo que algumas assumam, nesse contexto, uma acepção específica (por exemplo, *répondre à une question* [responder a uma questão]¹³).

<i>JOUER un rôle</i> [EXERCER um papel]
<i>AVOIR un effet</i> [TER um efeito]
<i>POSER un problème</i> [LEVANTAR um problema]
<i>OBTENIR des résultats</i> [OBTER resultados]
<i>AVOIR un impact</i> [TER um impacto]
<i>PRÉSENTER les résultats</i> [APRESENTAR os resultados]
<i>FAIRE l'hypothèse</i> [LEVANTAR a hipótese]
<i>PRENDRE la forme</i> [ASSUMIR a forma]
<i>ATTEINDRE l'objectif</i> [ATINGIR o objetivo]
<i>AVOIR une valeur</i> [TER um valor]
<i>DONNER des résultats</i> [FORNECER resultados]
<i>RÉPONDRE à une question</i> [RESPONDER a uma questão]
<i>UTILISER une méthode</i> [UTILIZAR um método]
<i>AVOIR une fonction</i> [TER uma função]
<i>AVOIR un rôle</i> [TER um papel]

Quadro 1: Algumas colocações frequentes em francês do tipo V-Nnos textos científicos.

O tratamento semântico e sintático desses dois tipos de expressões deve ser evidentemente diferenciado: assim como Mel'čuk em *Lexicologia Explicativa e Combinatória* (MEL'ČUK, 2011), acreditamos que os frasemas completos, não-composicionais, devem ser tratados como palavras simples. Eles serão, por exemplo, caracterizados com a ajuda de uma etiqueta semântica (*objet scientifique abstrait* [objeto científico abstrato], *verbe d'opinion* [verbo de opinião]), enquanto as colocações, composicionais, podem constar na base. As expressões *faire une hypothèse* [levantar uma hipótese], *réfuter une hypothèse* [refutar uma hipótese], *hypothèse valide* [hipótese válida] podem assim ser codificadas na entrada *hypothèse* [hipótese], já que cada elemento da colocação é objeto de um tratamento semântico (Cf. TUTIN, 2007b).¹⁴ No que concerne à extração automática das expressões, as técnicas divergem também de acordo com o tipo de sequências. As colocações, geralmente binárias, são frequentemente extraídas com a ajuda

¹³ Nesse contexto, *question* [questão] é quase sempre sinônimo de *problématique* [problemática], *problème* [problema], mais do que *phrase interrogative* [frase interrogativa].

¹⁴ O tratamento sintático desses elementos é igualmente diferenciado. A maior parte das colocações, enquanto sequências polilexicais composicionais, tem, na verdade, um comportamento sintático *normal* (por exemplo, *faire DET hypothèse* [levantar DET hipótese] se torna relativa e se reduz: *l'hypothèse qu'il a faite* [a hipótese levantada por ele], *l'hypothèse faite ici* [a hipótese elaborada aqui] etc.). Os frasemas completos são, em geral (mas nem sempre), menos variáveis no plano sintático.

de medidas estatísticas de correlação, que selecionam as expressões de duas palavras que mantêm uma relação sintática e cuja coocorrência é a mais significativa (SERETAN, 2011; DIWERSY e KRAIF, 2012). A situação mostra-se mais complicada para os frasemas completos, na medida em que a expressão pode ser composta por mais de dois elementos e corresponder a uma organização sintática menos previsível. Técnicas com base em bancos de árvores e medidas estatísticas¹⁵ apresentam, entretanto, resultados bastante encorajadores (CORMAN, 2012).

2.2. Sequências polilexicais de função discursiva (ou marcadores discursivos)

A noção de *marcador discursivo* recebe, de acordo com as escolas de pensamento, contornos bem flutuantes, e convém precisar qual sua acepção aqui. Nós os definimos como elementos linguísticos que servem para elencar segmentos discursivos, orações ou segmentos textuais maiores, para estruturar a argumentação e a organização textual (*dans un premier temps* [em um primeiro momento], *pour conclure* [para concluir], *en ce qui concerne* [no que concerne], *en d'autres termes* [em outras palavras], *par opposition à* [em oposição a], *c'est pourquoi* [é por isso que]...). Os marcadores discursivos não se limitam portanto aos conectores e podem ter também uma função metadiscursiva, por exemplo, uma operação de reformulação, como a desempenhada por *en bref* [em resumo], ou de paralelismo (*d'une part... d'autre part* [por um lado... por outro lado]) (Cf. PÉRY WOODLEY, 2000)¹⁶. Contrariamente a Siepmann (2007), não integraremos nessa classe as expressões de função interpessoal, como aquelas que exprimem a certeza, a dúvida, a necessidade e cuja função primeira não é articular segmentos textuais (*il ne fait aucun doute* [sem dúvida alguma], *il est probable* [é provável], *il apparaît indispensable* [parece indispensável] etc.), mas sim expressar uma modalização, um ponto de vista.

No texto científico, as sequências polilexicais de função discursiva cumprem várias funções: servem para guiar o locutor ao longo do texto (função metatextual) e para introduzir definições, exemplos, explicações, para expressar as relações lógicas e para estruturar o raciocínio, entre outras. Do ponto de vista sintático, esses elementos são basicamente adverbiais (em sentido amplo), locuções de função prepositiva ou conjuntiva, ou, até mesmo, sequências de base verbal (ex.: *il en découle que* [disso decorre que]). Siepmann (2007) observa acertadamente que essas expressões são muito mal estudadas e recenseadas, sobretudo no que tange aos marcadores de segundo nível, ou seja, os marcadores de frequência média¹⁷. Evidentemente essas sequências polilexicais de função

¹⁵ A medida utilizada é a Informação Mútua.

¹⁶ Seguindo o modelo de Charolles (2003), não integramos às sequências de função discursiva os *cadratifs* [N.T.: enquadres enunciativos conforme o uso de Ingedore Koch, 2006.] verídicos (ex.: *l'année dernière* [no ano passado] ou não verídicos (ex.: *à notre grand regret* [para nosso grande pesar])). Os primeiros serão considerados como sequências de função referencial, enquanto os segundos pertencem à categoria das sequências de função interpessoal (cf. seção 2.3).

¹⁷ Os marcadores de primeiro nível seriam, desse modo, expressões como *en effet* [com efeito], *en revanche* [em contrapartida] etc.

discursiva não são específicas para a maior parte dos textos científicos e podem ser encontradas em outros gêneros, como a vulgarização científica ou os textos jornalísticos. Será dada maior atenção às que são muito frequentes no gênero de nosso interesse¹⁸.

O tratamento linguístico dessas expressões não é trivial. Além de sua descrição semântica e discursiva, que parece complexa por sua ambiguidade, mostra-se essencial indicar seu modo de emprego sintático (por exemplo, a sequência aparece necessariamente no começo da frase? Sobre qual tipo de constituintes ela incide?). A extração dessas expressões, cujos elementos são adjacentes em geral, é mais simples do que a de frases completas. Uma simples extração de segmentos repetidos (Cf. LAFON e SALEM, 1983) a partir de um *corpus* lematizado, ainda que produza bastante *ruído*, é normalmente suficiente para a localização desses elementos¹⁹.

2.3. Sequências polilexicais de função interpessoal

Essas expressões são indissociáveis da situação de comunicação e remetem à dimensão interpessoal da linguagem. Consideramos, nessa categoria, as sequências polilexicais de função pragmática ou *pragmatemas*²⁰ (Cf. MEL'ČUK, 2011; ver também os *enunciados ligados* a uma situação de interação, conforme proposto por Fónagy, [1997]), esses enunciados autônomos – frases padrão ou frases sem verbo, por exemplo, *y'a pas de quoi* [não tem de quê]. *Je vous en prie* [Disponha]. *Après vous* [Você primeiro] –, inseparáveis de situações enunciativas específicas, sobretudo em um contexto de diálogo. Essas expressões são relativamente ausentes nos textos científicos fora do peritexto científico²¹, mas aparecem, com mais frequência, em contextos científicos orais²². Incluímos ainda, nas sequências polilexicais de função interpessoal, as sequências polilexicais de função modal, que servem para exprimir o ponto de vista do locutor quanto ao conteúdo oracional. Trata-se de expressões como *semble-t-il* [ao que parece]; *de mon point de vue* [de meu ponto de vista]; *selon toute probabilité* [provavelmente]; *il apparaît nécessaire de...* [parece ser necessário...]; *il est probable...* [é provável...], que, em nossa visão, não se classificam como marcadores discursivos na medida em que sua função primeira não é relacionar conteúdos de segmentos textuais, mas, muito mais, exprimir o ponto de vista do locutor sobre o conteúdo oracional apresentado. Pelo distanciamento que tomam do conteúdo oracional, essas sequências podem, de certa forma, ser analisadas como tendo uma função metadiscursiva, no sentido definido por Hyland (2005)²³. Do

¹⁸ Nesse caso, medidas estatísticas, como a especificidade, podem ser utilizadas.

¹⁹ A extração de base sintática tem a vantagem de gerar menos ruído, ou seja, menos respostas inapropriadas.

²⁰ A noção de pragmatema é mais restrita, visto que ela não considera necessariamente a polilexicalidade.

²¹ Agradecimentos redigidos com fórmulas estereotipadas, por exemplo.

²² Como nas interações orais de palestras. Por exemplo, o mediador após a fala de um palestrante: *any question?* [alguma pergunta?] ou *nous vous remercions encore pour cet exposé X* [agradecemos mais uma vez por essa apresentação X], em que X corresponde a um adjetivo avaliativo de polaridade positiva.

²³ Esse tipo de expressão polilexical assinala principalmente os primeiros e segundos tipos. Segundo Hyland: [Os princípios chave] são: 1. metadiscorso é distinto dos aspectos propositivos do discurso; 2. metadiscorso se

ponto de vista sintático, essas expressões são principalmente adverbiais (ex.: *selon toute apparence* [ao que tudo indica], *à première vue* [à primeira vista]), ou orações principais que introduzem o conteúdo oracional (*il est probable que* [é provável], *il est discutable de* [é discutível que] etc.). Como para as sequências discursivas, a extração dessas expressões, em que os elementos são muitas vezes adjacentes, pode ser realizada com auxílio da localização de segmentos repetidos.

3. Rotinas retóricas

Essas expressões se situam no limite da categoria fraseológica, na medida em que correspondem, por um lado, a enunciados completos e, por outro, comportam um léxico e estruturas sintáticas diversificadas. Trata-se de enunciados estereotipados que remetem a funções retóricas específicas ao gênero de nosso interesse, como a filiação científica indicada por um autor, a demarcação relativamente a um colega de área, a avaliação de atividades científicas de pares, a elaboração de provas factuais: por exemplo, *comme on peut le voir ci-dessous* [como se pode ver abaixo], *on l'a vu ci-dessus* [vimos acima]...

Essas rotinas estão mais associadas a uma função retórica (explicar, persuadir, avaliar, provar) do que a uma estrutura discursiva (estruturar o texto, indicar a relação entre os segmentos textuais); em contrapartida, elas põem em jogo a estrutura enunciativa do discurso por meio dos actantes em questão (por exemplo, o *je* [eu] autor do texto ou o *on* [sujeito indeterminado] que pode incluir o leitor testemunha) e utilizam um material lexical específico. Elas foram estudadas e descritas como padrões principalmente para o TAL no âmbito de trabalhos sobre a extração de informação, como veremos a seguir.

Estabelecemos padrões ou modelos – ou seja, modelizamos – através de esquemas abstratos, mais ou menos como os *frames* semânticos de Fillmore (2003). Mas, contrariamente a essa abordagem de inspiração semântica e cognitiva, esses esquemas têm aqui uma função retórica e são especificamente associados a esse gênero. Por exemplo, a frase *nous nous différencions de Schmigh (1997) sur ce point* [nós nos diferenciamos de Schmigh (1997) nesse ponto] pode ser analisada a partir de um esquema complexo em que entram em jogo um agente (*nous*), uma operação de *demarcação* (*différencions*), uma fonte (*Schmigh (1997)*), um objeto (*ce point*). As rotinas retóricas podem sem dúvida integrar sequências polilexicais mais reduzidas, em especial sequências polilexicais referenciais ou modais. Detalharemos esse modelo após esta seção.

Evidentemente essa tipologia de inspiração funcional não está isenta de problemas. Certos elementos podem se encaixar em várias categorias, e uma escolha deverá ser feita, em caso de ambiguidade, para as implementações do modelo. Por exemplo, as colocações que indicam uma avaliação (*modèle adapté* [modelo adaptado], *hypothèse valide* [hipótese válida], *résultat encourageant* [resultado encorajador]) correspondem, ao mesmo tempo, a expressões referenciais (dizem respeito a objetos específicos) e a sequências de função

refere a aspectos do texto que incorporam a interação escritor-leitor; 3. metadiscurso se refere apenas a relações que são internas ao discurso. (HYLAND, 2005, p.38).

interpessoal (expressam um ponto de vista). Para a classificação das expressões, privilegiaremos a estrutura sintática da sequência polilexical. Os exemplos mencionados, que têm uma estrutura colocacional, serão assim preferencialmente classificados como expressões referenciais. As etiquetas semânticas associadas às expressões poderão ser aproximadas de sequências polilexicais de função modal (ex.: *de mon point de vue* [de meu ponto de vista]).

O Quadro 2 sintetiza a tipologia proposta.

Tipo	Definição	Subtipo	Definição	Tipos Sintáticos	Exemplos no Texto Científico
Sequências polilexicais de função referencial	Expressões recorrentes que remetem a noções no texto científico	Frasemas completos (ou expressão fixa)	Expressões não-composicionais	Principais categorias sintáticas abertas (adjetivo, nome, verbo principalmente)	<i>Tirer parti</i> [tirar partido] <i>Prendre en compte</i> [levar em conta] <i>Point de vue</i> [ponto de vista]
		Colocações lexicais	Expressões recorrentes, composicionais, geralmente binárias.	Associações sintáticas de tipo <i>Nom-suj-Verbe</i> [N-suj-V]; <i>Verbe-obj-N</i> [V-obj-N.]; <i>Adj-Epit-Nom</i> [Adj-adj-N]; <i>Nom-Prep-Nom</i> [N-Prep-N]	<i>Faire une hypothèse</i> [levantar uma hipótese] <i>Résultats prometteurs</i> [resultados promissores] <i>Hypothèse de départ</i> [hipótese de partida]
Sequências polilexicais de função discursiva	Expressões recorrentes que servem para estruturar o discurso.			Sequências adverbiais, prepositivas, conjuntivas ou construídas em torno de uma base verbal.	<i>Dans un premier temps</i> [em um primeiro momento] <i>Il en va de même</i> [é o mesmo caso de]...

Sequências polilexicais de função interpessoal	Expressões relacionadas à situação de enunciação.	Sequências polilexicais de função pragmática.	Sequências estereotipadas associadas a uma situação de enunciação específica.	Enunciados geralmente autônomos (frases padronizadas ou frases sem verbo) (quase ausentes do discurso científico escrito).	No campo de agradecimento nos artigos: <i>Un grand merci à mes collègues X et Y pour leur relecture</i> Adj. [muito obrigado aos meus colegas X e Y por sua leitura Adj.]
		Sequências polilexicais de função modal	Introduzem um ponto de vista sobre o conteúdo oracional.	Elementos adverbiais ou que introduzem orações (completiva ou infinitiva).	<i>En quelque sorte</i> [de certa forma] <i>Il est nécessaire de</i> [é necessário]...
Rotinas retóricas	Enunciados típicos do texto científico que correspondem a uma função retórica específica.			Estruturas frásticas às quais se pode associar um esquema abstrato.	<i>Nous reprenons ici le modèle développé par X</i> [retomamos aqui o modelo desenvolvido por X] <i>Nous nous distinguons de X sur ce point</i> [distinguímo-nos de X nesse ponto]

Quadro 2: Tipologia de sequências polilexicais no texto científico

3.1. Modelos de rotinas retóricas²⁴

O texto científico apresenta inúmeros níveis de pré-fabricação linguística, não apenas no que concerne às sequências lexicais e à estrutura discursiva, muitas vezes limitada, mas também na formulação de certas funções retóricas. Essas formulações estereotipadas dos textos científicos, que chamamos de *rotinas retóricas*, situam-se no limite da fraseologia: se correspondem, por um lado, a um tipo de *prêt-à-écrire* do gênero científico e dão frequentemente a impressão de *já terem sido lidas*, também correspondem, por outro, a enunciados completos, construídos em torno de um predicado, e evidenciam

²⁴ Esta seção retoma, em parte, um capítulo de uma obra (CF. TUTIN, 2010B).

um léxico variado. Mais complexas no plano semântico e mais variadas no plano lexical, diferenciam-se das sequências polilexicais anteriores. Eis alguns exemplos dessas rotinas:

- Fornecer uma prova com a ajuda de um fato: *comme nous le voyons sur ce tableau...*[como podemos ver neste quadro...] *nous pouvons observer que...*[podemos observar que...];
- Contraste e comparação: *Contrairement à Parker (1990), nous...* [Contrariamente a Parker (1990), nós...], *Notre étude diffère de Parker (1990)...* [Nosso estudo difere de Parker (1990)...];
- Filiação científica e acadêmica: *À la suite de Parker (1970), nous...* [Dando seguimento a Parker (1970), nós...], *Nous reprenons la définition de Parker (1979)* [Retomamos a definição de Parker (1979)], *Notre modèle reprend les travaux de Parker (2010)* [Nosso modelo retoma os estudos de Parker (2010)];
- Definir uma problemática: *Notre article traite de la phraséologie scientifique* [nosso artigo trata da fraseologia científica]; *l'objet de notre article est la phraséologie scientifique* [o objeto de nosso artigo é a fraseologia científica].

3.2. Rotinas retóricas e modelos da Linguística e do Processamento Automático de Linguagem

As rotinas apresentam uma grande variedade lexical e sintática, no entanto, correspondem a um esquema semântico e retórico recorrente em um tipo de texto. Elas foram, até onde sabemos, bem pouco estudadas pela Linguística, porém se aproximam de muitas noções desenvolvidas no âmbito da fraseologia: têm pontos em comum com a noção de *motif*²⁵ criada por Mellet e Longrée, que a aplicaram ao estudo dos textos em latim (ver, por exemplo, Mellet e Longrée, 2012). O *motif* é definido como uma estrutura recorrente e linear, com uma função textual e discursiva. Nossa definição das rotinas é, ao mesmo tempo, mais estrita, na medida em que a função das rotinas é sobretudo *retórica* (a função discursiva caracterizará outros tipos de sequências, como já apresentado), e mais ampla, dado que a linearidade da expressão não nos parece uma característica chave. Exemplificando, as sequências *Contrairement à Parker (1990), nous...* [contrariamente a Parker (1990), nós...] e *Notre étude diffère de Parker (1990)* [nosso estudo difere do de Parker (1990)] recorrem a uma organização de actantes diferente (a fonte antes do agente no primeiro exemplo e de forma inversa no segundo), entretanto correspondem ao mesmo esquema para nós. Em contrapartida, parece essencial que os diferentes elementos presentes nessas estruturas estejam relacionados semântica e sintaticamente. Além disso, a noção de rotina retórica é bastante próxima daquela de *estruturas pré-fabricadas*, em particular da noção de *modelo de texto* desenvolvida por Gülich e Krafft (1997):

²⁵ N.T.: Do francês, *motif*: “[...] associação recorrente de X elementos do texto, seguindo uma determinada estrutura sintagmática, com função textual de encadeamento da informação.” (Sylvie Mellet, Dominique Longrée. *Légitimité d’une unité textométrique: le motif*. Anne Dister; Gérald Purnelle; Dominique Longrée. JADT 2012, 11èmes Journées internationales d’analyse statistique des données textuelles, Jun 2012, Liège, Belgique. Revue en ligne Lexicometrica, p. 717, 2012.).

Nomeamos *modelos de textos* estruturas pré-fabricadas de que nos servimos para a produção de certos tipos de texto [...]. Do ponto de vista da produção, o modelo é uma forma de receita que comporta indicações sobre as condições de emprego do texto: Onde? Em que momento? Em que condições?; Sobre sua estrutura: quais tipos de informações, eventualmente, em que ordem? Há prioridades a serem respeitadas?; e, com frequência, listas de expressões para utilização referencial (GÜLICH e KRAFFT, 1997, p. 258).

A noção de modelo de texto parece, contudo, mais global. As rotinas que descrevemos não se inserem necessariamente em uma estrutura linear: são enunciados estereotipados aos quais o locutor pode recorrer em qualquer parte do texto, mesmo que algumas funções sejam especialmente mais frequentes em certas partes, como a introdução. Por exemplo, as rotinas de prova científica (ex.: *comme on le voit dans le tableau 3* [como se vê no quadro 3]) são encontradas principalmente na parte metodológica, mas raramente nas introduções ou nas conclusões.

Se esse tipo de enunciado estereotipado foi, por fim, muito pouco estudado pela linguística, o TAL, sobretudo a área científica de extração de informação, propôs modelos mais sólidos. Esse domínio do TAL visa localizar, nos textos científicos, as partes que correspondem a conteúdos semânticos específicos. O estabelecimento de padrões de bases retórica e pragmática desenvolvidos a partir do TAL parecem mais estruturados (cf. Teufel, 1998; Sándor, 2007). O modelo de Teufel (1998), o AZ (*Argumentative Zoning*), inicialmente utilizado para o resumo automático, é ainda uma referência de TAL. Exemplificando, a expressão *to our knowledge* [até onde sabemos] é associada à função *gap introduction*; por sua vez, *when compared to* [quando comparado a] é associada à função de *comparação*. Porém, como observou Sándor (2007), as expressões levantadas por Teufel são apenas *pacotes lexicais* não organizados do ponto de vista sintático ou semântico: não são padrões fraseológicos, mas apenas listas de palavras, sem organização específica. Para preencher essa lacuna, Sándor propõe um modelo alternativo, o *concept matching*, que inclui uma representação sintática e semântica combinada com uma função discursiva, inspirada em parte por Swales (1990) (CARS): um conjunto de conceitos, realizados pelos marcadores lexicais e ligados pelas relações sintáticas, é associado a uma função retórica. Por exemplo, a função *conhecimento partilhado* integra os conceitos *passado, geral, pesquisador, atividade mental, objeto/resultado da atividade científica*, em que cada um se realiza de forma diferente no plano lexical. O Quadro 3, extraído de Sándor (2007), oferece uma versão esquematizada desse exemplo.

<i>background</i> [compartilhado]		<i>Knowledge</i> [conhecimento]		
<i>general</i> [geral]	<i>past</i> [passado]	<i>researcher</i> [pesquisador]	<i>mental activity</i> [atividade mental]	object result of research activity [objeto/resultado da atividade científica]
<i>most</i> [a maior parte]		<i>evolutionists</i> [evolucionistas]	<i>Consider</i> [considerar]	
	<i>recent</i> [recente]		<i>get a picture</i> [entender]	<i>findings</i> [descobertas]

Quadro 3: Exemplo da função discursiva conhecimento compartilhado (Sándor, 2007).

A abordagem proposta por Sándor parece de fato interessante, na medida em que sistematiza o léxico e as funções semânticas associadas à função retórica. Retomamos alguns de seus princípios ao sistematizar, por um lado, os papéis semânticos e, por outro, ao considerar os aspectos retóricos e enunciativos.

3.3. Frames semânticos

Propomos um modelo de representação, os *frames* semânticos, que permite sistematizar os enunciados estereotipados que correspondem a uma função retórica. Esses *frames* combinam as funções pragmáticas e retóricas com a dimensão semântico-enunciativa, assim como com aspectos lexicais e sintáticos (Cf. TUTIN, 2010b). A representação inclui três níveis:

- O nível retórico depende fortemente das estratégias discursivas em uso no gênero textual envolvido. No texto científico, é possível identificar, por exemplo, funções como *a demarcação em relação aos pares*, *a apresentação da problemática de pesquisa*, *a filiação científica*. É sem dúvida difícil propor aqui uma longa lista dessas funções, que se estabelece empiricamente através do estudo de *corpus*.²⁶
- O nível semântico e enunciativo estabelece padrões com base nos esquemas abstratos associados às funções retóricas. Por exemplo, o *frame* da filiação científica articula um autor científico que toma de empréstimo um objeto científico desenvolvido por um autor fonte. Esse *frame* inclui os tipos semânticos de predicados, os papéis semânticos (no sentido de Fillmore) e os tipos semânticos dos participantes da situação enunciativa. Os tipos semânticos dos participantes são estreitamente ligados ao gênero e remetem aos protagonistas de um texto científico (o autor do texto, o leitor, a comunidade científica). Um *frame* conforme o de Fillmore (ex.: Fillmore *et al.*, 2003) permite estabelecer um padrão nesse nível. Entretanto, contrariamente aos *frames* semânticos do autor, nosso modelo não é cognitivo, mas sim semântico e textual: propõe uma representação abstrata de estereótipos textuais. Além disso, os tipos e papéis semânticos são estreitamente ligados ao gênero e só têm validade se associados a um modelo textual.

²⁶ Este trabalho está sendo realizado por nossa equipe.

- O nível sintático e lexical inclui os marcadores lexicais associados aos tipos semânticos, já as relações sintáticas, a seus papéis semânticos. Essa correspondência elucida a variedade lexical e sintática dos estereótipos textuais.

Tomemos o exemplo da filiação científica (Cf. GROSSMANN *et al.*, 2009), que pode ser considerada como uma função retórica clássica ligada ao texto científico, por meio da qual o autor indica sua pertença a uma linha, sua intenção de continuar um modelo teórico ou que faz uso de um modelo para o propósito de um estudo. Os exemplos seguintes, extraídos de textos científicos, dizem claramente respeito a essa noção de filiação:

- (1) *Cette étude reprend ici les travaux de Bunt (1995) sur...* [Este estudo retoma os trabalhos de Bunt (1995) sobre...]
- (2) *À la suite de (Lee 1986), nous ne considérerons pas que l'employeur...* [Dando seguimento a (Lee 1986), não consideraremos que o empregador...]
- (3) *Nous avons choisi d'utiliser le modèle des graphes conceptuels (Sowa, 1984)...* [Escolhemos utilizar o modelo de gráficos conceituais (Sowa, 1984)...]

Apesar da diversidade das realizações lexicais e sintáticas, um esquema comum da filiação pode ser associado a esses exemplos. Essa função retórica articula três participantes e um predicado genérico de *empréstimo*.

- um autor científico, em geral o autor do texto (*nous* [nós], *je* [eu]) ou, por metonímia, o objeto que ele constrói no texto (*cette étude* [este estudo], *ce travail* [este trabalho], *notre étude*²⁷ [nosso estudo]). Esse participante nem sempre se realiza na superfície por um marcador explícito. Por exemplo, em *le modèle reprend la notion de cadre développée par Fillmore* [o modelo retoma a noção de *frame* desenvolvida por Fillmore], não é possível, fora do contexto, determinar se é o modelo do autor ou de um terceiro.
- o objeto científico (*travaux* [estudos], *modèle* [modelo], mas também *terme* [termo], *idée* [ideia], *expression...* [expressão...]) é tomado emprestado ou utilizado pelo autor; trata-se das construções intelectuais que circulam na comunidade científica. O paradigma dos *objets scientifiques* é bastante amplo e depende evidentemente da disciplina. Em Grossmann et al. (2009), identificamos que os objetos científicos utilizados em Economia eram, com frequência, mais complexos (*analyse* [análise], *approche* [abordagem], *modèle* [modelo]) do que aqueles que apareciam em Linguística (*termes* [termos], *notion* [noção], *concept* [conceito]). Há casos em que a filiação não menciona especificamente o objeto utilizado, como no exemplo (2) acima: nesse caso, é a análise global do autor que é retomada.
- a fonte do empréstimo é uma referência, geralmente um autor ou autores associados a uma data (*Bunt (1995)*, [*Sowa, 1984*], grifos da autora), sendo

²⁷ Um dêitico de lugar pode também ser usado para remeter a um autor: *Le modèle développé ici reprend...* [O modelo desenvolvido aqui retoma...].

esta a convenção nos textos científicos. Em alguns casos, a fonte não se realiza pela convenção autor-data, mas por um adjetivo derivado (ex.: *nous reprenons la notion fillmorienne de cadre* [retomamos a noção fillmoriana de *frame*]).

- finalmente, um predicado verbal (*reprendre* [retomar], *utiliser* [utilizar]) ou nominal integrado a um sintagma preposicional (*à la suite de* [dando seguimento a], *dans la continuité de* [dando continuidade a]) indica o processo de empréstimo científico.

Os papéis semânticos (aqui chamados de maneira convencional *agente* [agente], *objet* [objeto] e *source* [fonte]) ligam o predicado “Empréstimo” aos objetos semânticos. A Figura 1 esquematiza essa análise.

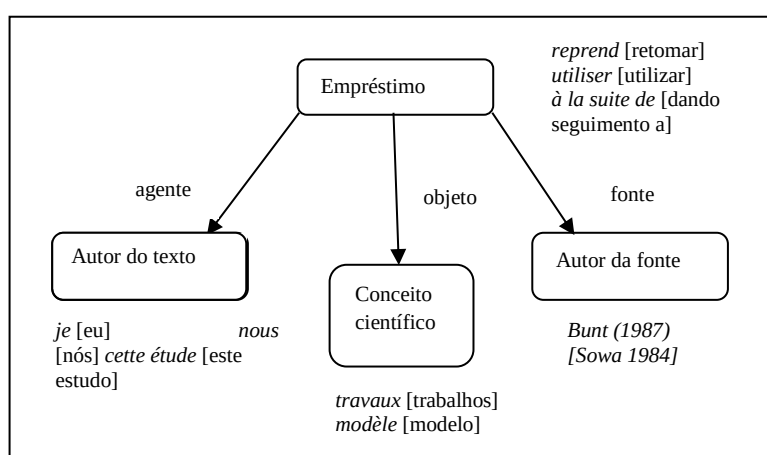


Figura 1: *Frame* semântico-retórico da filiação científica.

Do ponto de vista sintático-semântico, várias construções podem ser associadas a esse *frame*, como se observa na Figura 2. Essas construções são aqui representadas com a ajuda de árvores de dependência, do mesmo tipo que aquelas utilizadas na análise sintática para o projeto Scientext.

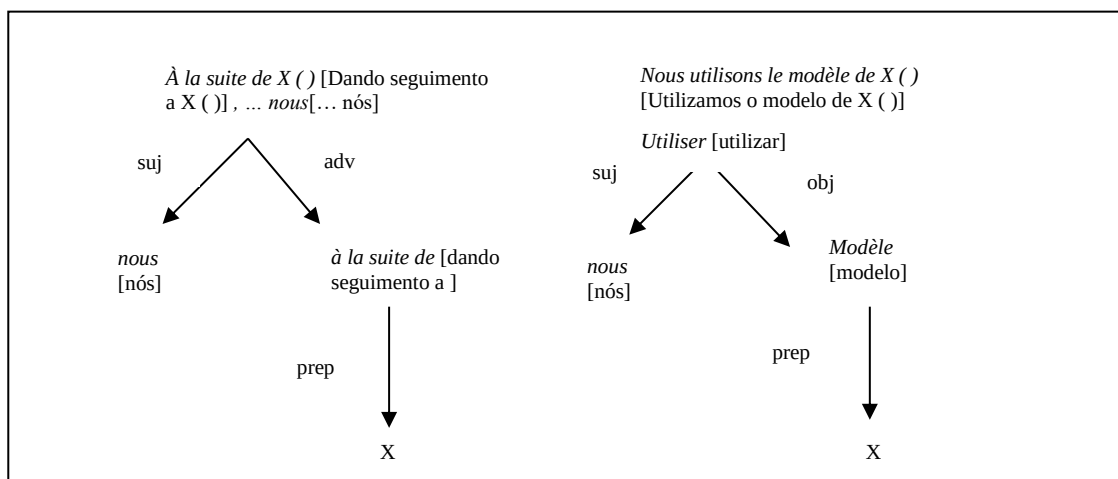


Figura 2: Exemplos de configurações sintáticas que correspondem ao *frame* da filiação científica.

Em frases como *à la suite de X...* [dando seguimento a X...], o conceito de *empréstimo* é introduzido pela preposição, enquanto em *nous utilisons le modèle de X* [utilizamos o modelo de X], o autor é o sujeito, e o conceito de *empréstimo* está ligado a um verbo. No projeto Scientext, cerca de 15 gramáticas locais foram implementadas de maneira simplificada seguindo esse princípio, as diferentes configurações sintáticas utilizam quase-sinônimos que remetem ao mesmo objeto semântico. Algumas relações sintáticas de dependência estão associadas a uma relação semântica: por exemplo, as construções ativas (ex.: *formuler une hypothèse* [levantar uma hipótese]), as construções passivas (ex.: *l'hypothèse est formulée* [a hipótese é levantada]) ou as passivas reduzidas (ex.: *hypothèse formulée* [hipótese levantada]) corresponderão a uma relação da ordem do “objeto profundo”, como proposto em Hagège e Roux (2003). Contudo, por razões de eficácia informática, versões simplificadas dessas gramáticas são hoje utilizadas no site Scientext²⁸. As gramáticas baseiam-se na sintaxe do *corpus* analisado com ajuda do *software* Syntex, desenvolvido por Bourigault (2007). Essa linguagem de gramáticas locais, Conquest, foi desenvolvida por Kraif (2008) e formatada para a interface do Scientext e do Scienquest por Falaise (2013).

A Figura 3 apresenta uma versão simplificada da gramática da filiação. Observa-se que o autor pode ser um pronome de primeira pessoa ou *on* [sujeito indeterminado]. O mecanismo de empréstimo se realiza através de um verbo como *utiliser* [utilizar], *reprendre* [retomar], *choisir* [escolher], enquanto o objeto utilizado é um nome que pertence a uma lista de construtos científicos como *analyse* [análise], *méthode* [método] ou *approche* [abordagem]. A gramática elaborada é aplicada no *corpus* Scientext e permite extrair um conjunto de rotinas que exprimem a filiação.

```
// Frame da filiação
// Objeto
$Objet [$Objeto] = analyse [análise], approche [abordagem], concept [conceito], courant [corrente], critère
[critério], définition [definição], démarche [andamento], démonstration [demonstração], dénomination
[denominação], distinction [distinção], étude [estudo], facteur [fator], hypothèse [hipótese], mesure
[medida], méthode [método], modle [modelo], modélisation [modelização], notion [noção], prncipe
[princípio], procédure [procedimento], proposition [proposta], recherche [pesquisa], règle [regra], référence
[referência], remarque [observação], réflexion [reflexão], représentation [representação], statistique
[estatística], technique [técnica], terme [termo], terminologie [terminologia], théorème [teorema], théorie
[teoria], travail [trabalho], typologie [tipologia]
// Autor
$Auteur[$Autor] = nous[nós], je[eu], on[]
// Verbo de empréstimo
$Vemprunt [$Vempréstimo] =utiliser [utilizar], exploiter [explorar], employer [empregar], retenir [manter],
choisir [escolher], compléter [completar], poursuivre [prosseguir], prolonger [prolongar], reprendre
[retomar], retenir [manter], étendre [ampliar], compléter [completer], suivre [seguir], Inscrire [inscrever],
recourir [recorrer]
Controle      =      <lemma=$Auteur,#1>&&<lemma=$Vemprunt,#2>&&<lemma=$Objet,#3>      ::
(SUJ,#2,#1)(OBJ,#2,#3)
```

Figura 3: Uma gramática simplificada que representa o *frame* da filiação teórica.

²⁸ Objetivamos acrescentar diretamente ao *corpus* analisado essas relações semânticas de maneira a acelerar o processo de consultas no *site* que, neste momento, é muito demorado.

A gramática elaborada não é evidentemente perfeita – gera ruído e silêncio²⁹ –, mas permite extrair um conjunto de contextos pertinentes que ilustram esse tipo de rotinas, como os exemplos abaixo.

- (4) *Nous reprenons ici partiellement la typologie proposée dans Segond & Breidt (1995)...* [Retomamos parcialmente aqui a tipologia proposta em Segond & Breidt (1995)...] (Linguística – comunicação)
- (5) *Ce dernier potentiel ne permettant pas d'accéder à l'ensemble des énergies souhaitées, nous utiliserons donc la procédure de l'état de transition de Slater combiné au potentiel BP86...* [Como este último potencial não permite acessar o conjunto de energias desejadas, utilizaremos então o procedimento do estado de transição de Slater combinado ao potencial BP86...] (Eletrônica – tese)
- (6) *Nous utilisons le modèle des patrons (Blache et Rauzy, 2006) pour calculer la probabilité de chaque entrée du lexique global compte tenu du contexte syntaxique constitué par la séquence des formes orthographiques déjà saisies dans la phrase.* [Utilizamos o modelo de padrões (Blanche e Rauzy, 2006) para calcular a probabilidade de cada entrada do léxico global dado o contexto sintático constituído pela sequência de formas ortográficas já indicadas na frase.] (Processamento automático de linguagem – artigo)
- (7) *Nous retiendrons la règle de l'épenthèse (Klausenberger, 1974 ; Tranel, 1974), car elle est encore aujourd'hui envisagée, cependant sous une forme différente (Côté, 2005)* [Nós manteremos a regra da epêntese (Klausenberger, 1974 ; Tranel, 1974), pois ela é ainda hoje considerada, mas de uma forma diferente (Côté, 2005)] (Linguística – tese)

Futuramente desejamos melhorar a implementação desses *frames*, que é ainda muito rudimentar para dar conta da diversidade dos padrões, desenvolvendo a análise semântica acima. Desejamos, da mesma forma, desenvolver o estudo das funções retóricas de maneira a aumentar a apreensão dos fenômenos. Esse trabalho não pode, entretanto, ser automatizado, contrariamente aos outros tipos de sequências lexicalizadas que podem ser extraídas semiautomaticamente, e requer um profundo conhecimento do texto científico. Por fim, em colaboração com nossos colegas da didática e especialistas no processamento automático dos *corpora*, desejamos refletir sobre as aplicações didáticas simples que utilizem os resultados dos recursos que poderão ser integrados às atividades de observação e aos exercícios.

5. Conclusão

A fraseologia transdisciplinar dos textos científicos é diversa. A abordagem que propusemos aqui é essencialmente funcional e distingue quatro grandes tipos de sequências polilexicais segundo sua função predominante: (a) colocações e expressões fixas

²⁹ O ruído corresponde às extrações não apropriadas, já o silêncio concerne aos exemplos presentes no texto que não são detectados pelo recurso.

de função referencial que remetem primordialmente aos procedimentos e aos objetos científicos; (b) marcadores discursivos que estruturam o discurso ou a argumentação; (c) sequências polilexicais de função interpessoal que remetem aos traços da situação enunciativa e à presença do locutor/autor; (d) rotinas retóricas que remetem a funções retóricas estereotipadas do gênero em questão. É provável que a tipologia esboçada aqui não esgote todas as possibilidades, além disso, sua aplicação em um âmbito descritivo, o que estamos em vias de realizar, certamente não se dará sem dificuldades³⁰.

Voltamo-nos, em seguida, para as rotinas retóricas que, pela diversidade de suas realizações lexicais, situam-se no limite da fraseologia, mas apontam para a retórica estereotipada dos textos científicos. O estabelecimento de padrões proposto, inspirado livremente nos *frames* semânticos de Fillmore, foi implementado de forma simplificada com a ajuda de gramáticas locais e precisa ainda ser validado com um número maior de dados. Por último, ainda que o modelo tenha sido aplicado, até o momento, principalmente aos textos científicos – em que as diferentes funções retóricas se realizam em uma língua fortemente rotinizada –, ele pode ser, ao que tudo indica, utilizado para a descrição de outros socioletos, como as línguas profissionais ou o discurso político.

Na continuação do projeto Scientext, desejamos desenvolver os recursos fraseológicos conforme a tipologia proposta para o estudo epistemológico dos textos científicos e para aplicações didáticas.

Referências

- BIBER D., *University Language: a corpus-based study of spoken and written registers*, Amsterdam/Philadelphia, John Benjamins, 2006.
- BOURIGAULT D., *Un analyseur syntaxique opérationnel : SYNTEX*, mémoire de HDR, Toulouse, 2007.
- BURGER H., *Phraseologie: eine Einführung am Beispiel des Deutschen*, E. Schmidt, 2007.
- CAVALLA C., « Les écrits universitaires des étudiants étrangers : quelles normes présenter ? », BERTRAND O., SCHAFFNER I. (dir.), *Quel français enseigner ? La question de la norme dans l'apprentissage/enseignement*, Paris, Editions Ecole Polytechnique, 2010, p. 203-214.
- CORMAN J., *Extraction d'expressions polylexicales sur corpus arboré*, Mémoire de master 2 recherche, Sciences du langage, Grenoble, Université Grenoble 3 – Stendhal, 2012.
- COXHEAD A., “The Academic Word List: A Corpus-based Word List for Academic Purposes”, KETTERMAN B., MARKS G. (ed), *Teaching and Language Corpora (TALC) 2000 Conference Proceedings*, Atlanta, Rodopi, 2002, p. 73-89.

³⁰ Em colaboração com Thi Thu Hoai Tran, doutorando, e no âmbito do projeto ANR-Contint Termith com Marie-Paule Jacques, Cristelle Cavalla e Sylvain Hatier.

DIWERSY S., KRAIF O. « Le Lexicoscope : un outil pour l'analyse de profils combinatoires et l'extraction de constructions lexico-syntaxiques », *Actes de la conférence conjointe JEP-TALN-RECITAL 2012*, volume 2, juin 2012, Grenoble, p. 399-406.

FALAISE, A. « Exploitation linguistique de corpus arborés d'écrits scientifiques à l'aide du logiciel ScienQuest », TUTIN, A. ; GROSSMANN, F. (dir.), *L'écrit scientifique : du lexique au discours*, Rennes : PUR, 2013, p. 123-141.

FILLMORE C. J., JOHNSON C. R., PETRUCK M. R. "Background to Framenet", *International Journal of Lexicography*, n°16(3), 2003, p.235–250.

FLØTTUM K., DAHL T., KINN T. *Academic Voices. Across languages and disciplines*, Amsterdam/Philadelphia, John Benjamins, 2006.

FÓNAGY I., « Figement et changement sémantiques », M. MARTINS-BALTAR (ed), *La Locution entre langue et usages*, Fontenay-Saint-Cloud : ENS Editions, 1997, p. 131-164.

GLEDHILL C., *Collocations in science writing*, Language in performance, 22, Tübingen, Gunter, 2000.

GRANGER S., PAQUOT M., "Disentangling the phraseological web", Granger S., Meunier F, *Phraseology: An interdisciplinary perspective*, Amsterdam/Philadelphia, John Benjamins, 2008.

GROSSMANN F., TUTIN A. (dir.), *Les collocations : analyse et traitement*, Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, 2003.

GROSSMANN F., TUTIN A. GARCIA DA SILVA P., « Filiation et transferts d'objets scientifiques dans les écrits de recherche », *Pratiques*, n°143-144, 2009, p. 187-202.

GÜLICH E., KRAFFT U., « Le rôle du préfabriqué dans les processus de la production discursive », MARTINS-BALTAR M. (dir.), *La locution entre langue et usages*, Paris, ENS éditions. Fontenay-St Cloud, 1997, p. 241-276.

HAGEGE C., ROUX C. « Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes », *Actes de TALN, TALN 2003*, Batz-sur-Mer, 11-14 juin 2003.

HALLIDAY M.A.K., *Introduction to functional grammar*, 2nd ed., London, Edward Arnold, 1994.

HOWARTH P.A., *Phraseology in English Academic Writing*. Tübingen, Max Niemeyer, 1996.

HYLAND K., *Disciplinary discourses: social interactions in academic writing*, London, Longman, 2000.

HYLAND K., *Metadiscourse. Exploring Interaction in Writing*, London, Continuum, 2005

KOCOUREK R., *La langue française de la technique et de la science : vers une linguistique de la langue savante*, Wiesbaden, O. Brandstetter, 1991.

KRAIF O., « Comment allier la puissance du TAL et la simplicité d'utilisation ? l'exemple du concordancier bilingue ConcQuest », *Actes des 9ème Journées d'analyse statistique des données textuelles, JADT 2008*, Lyon, Presses universitaires de Lyon, 2008, p. 625-634.

LAFON P., SALEM A. « L'Inventaire des segments répétés d'un texte », *Mots*, n°6, 1983, p.161-177.

MANGIANTE J.-M., PARPETTE C., *Le français sur objectif universitaire*, Grenoble, Presses Universitaires de Grenoble, 2011.

MEL'CUK I. « Phrasèmes dans le dictionnaire », MEJRI S., ANSCOMBRE J.CL (dir.), *Le figement linguistique : la parole entravée*, Honoré Champion, Paris, 2011, p. 41-61.

MELLET S., LONGREE D., « Légitimité d'une unité textométrique : le motif », *Lexicometrica, Actes de JADT 2012*, 2012, p. 712-728.

PECMAN M., *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*, Thèse de doctorat, Université de Nice Sophia Antipolis, 2004.

PECMAN M., « Approche onomasiologique de la langue scientifique générale », *Revue Française de Linguistique Appliquée*, Vol XII, 2007-2, p. 79-96.

PERY-WOODLEY M.-P., *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*, Mémoire d'habilitation à diriger des recherches, Carnet de grammaire, Rapport n° 8, 2000.

SÁNDOR Á., "Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts", *Revue Française de Linguistique Appliquée*, XII-2, 2007, p. 97-109.

SERETAN V., *Syntax-Based Collocation Extraction*, Dordrecht, Springer, 2011.

SIEPMANN D., « Les marqueurs de discours polylexicaux en français scientifique », *Revue française de linguistique appliquée*, 2/2007, (Vol. XII), p. 123-136.

SWALES J. M., *Genre Analysis: English in academic and research settings*, Cambridge, Cambridge University Press, 1990.

TEUFEL S. « Meta-discourse markers and problem-structuring in scientific articles », *ACL 1998 Workshop, Discourse Structure and Discourse Markers*. Montreal, Somerset/New Jersey, 1998, p. 43-49.

TUTIN A. [2007a] (dir.), *Revue Française de Linguistique Appliquée, volume XII-2, Lexique et écrits scientifiques*, 2010.

TUTIN A. [2007b], « Collocations du lexique transdisciplinaire des écrits scientifiques :

annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique », *Cahiers de l'Institut de linguistique de Louvain*, n°Vol. 31, n° 2-4, 2007, p. 247-262.

TUTIN A. [2010b], "Showing phraseology in context: an onomasiological access to lexico-grammatical patterns in corpora of French scientific writings", Granger S., Paquot M. (ed), *Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, Cahiers du Cental, Louvain, Presses universitaires de Louvain, 2010, p. 302-312.

Como citar este texto (ABNT):

TUTIN, A. Tradução de Anna Carolina Teles. A fraseologia transdisciplinar dos textos científicos: das colocações às rotinas retóricas. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 170-190, 2018.