DICIONÁRIOS SENSÍVEIS AO CONTEXTO E INTEGRADOS A ASSISTENTES DE ESCRITA EM L2: NOVOS DESAFIOS PARA A LEXICOGRAFIA[1]

Sven Tarp[2]
Kasper Fisker[3]
Peter Sepstrup[4]
Tradução: Guilherme S. de Oliveira[5]
Supervisão e revisão: Márcia Moura da Silva[6]

"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 Africânder 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] Bacharelando 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 Bergenholtz *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 Kilgarriff (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:

"*to a beautiful work*" (com um belo trabalho)
"*to a beautiful woman*" (com uma bela mulher)
"*to a beautiful world*" (com um belo mundo)
"*to a beautiful word*" (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.

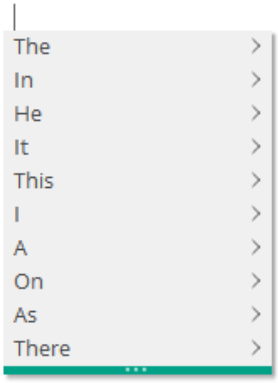
Como foi o caso acima, cada uma dessas tuplas de quatro palavras receberá uma nota do modelo de língua. Felizmente, "*The man was married*" é uma das tuplas de maior nota, o que sugere que *married* é a tradução mais provável para *gift* nesse contexto real. Uma vez que uma nota foi dada, os candidatos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

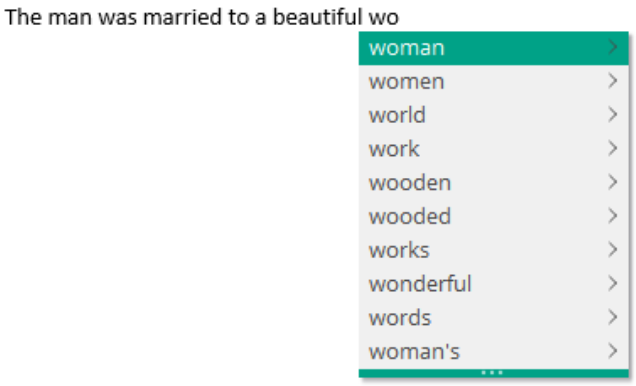
Figura 1 – Palavras iniciais



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.

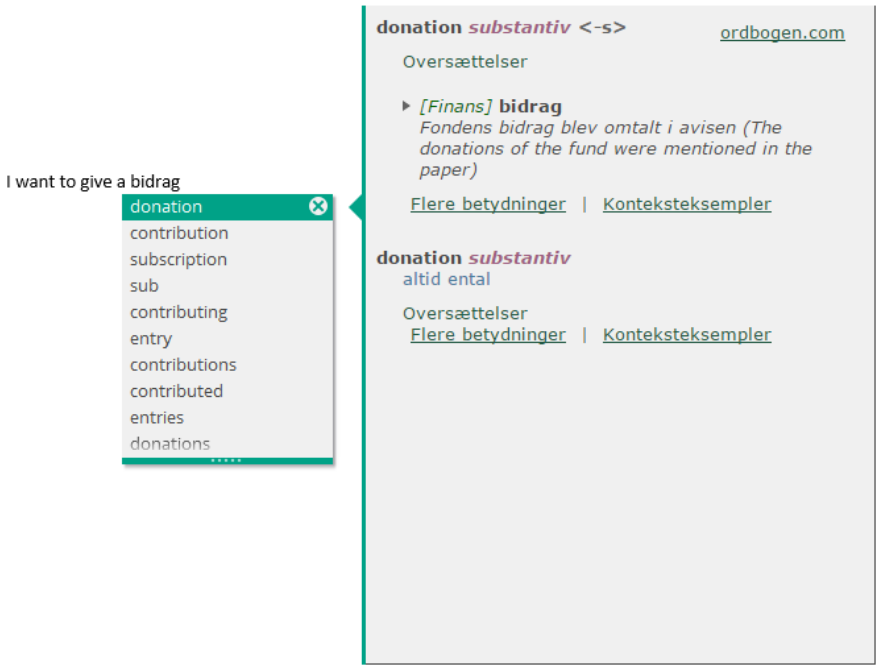
Figura 2 – A tupla "*to be a beautiful wo*" + finalização



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

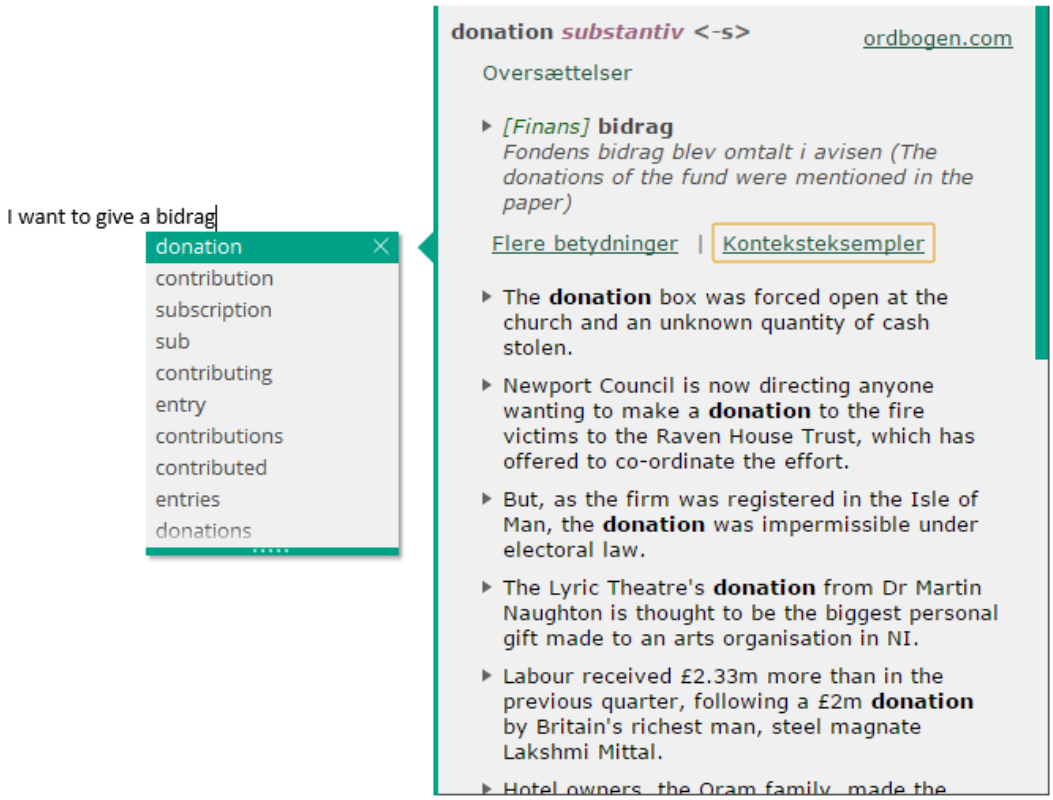
A entrada padrão do dicionário reproduzida na Figura 3 fornece apenas um sentido de *donation*. Apesar disso, a janela oferece várias opções ao usuário, que pode clicar em *Flere betydninger* (Mais definições) ou *Kontekseksempler* (Exemplos de contexto) para mais dados (ou até no *ordbogen.com* para 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 *Kontekseksempler*. 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 *Kontekseksempler* 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).

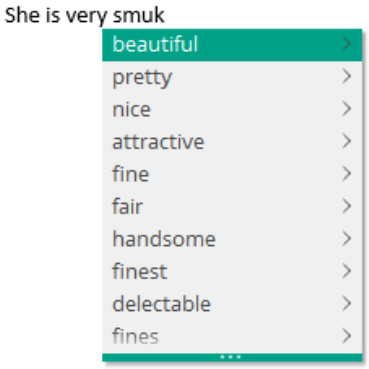
Figura 5 – A tupla "*to give a bidrag*" + exemplos de sentença



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).

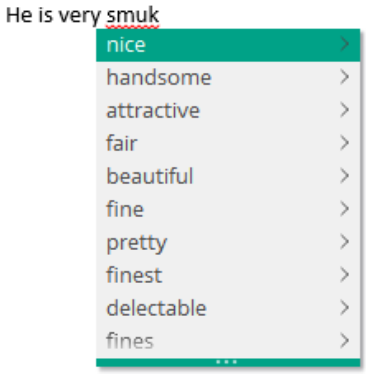
Figura 6 – A tupla "*She is very smuk*" + equivalentes



Fonte: Acervo do autor

Entretanto, se o autor, em vez disso, quiser 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 *Kontekstеksempler*.

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.

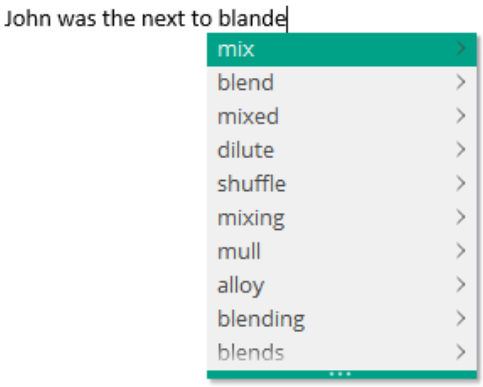
2) Quando um autor não sabe a palavra a ser usada em inglês, ele precisa de uma solução L1-L2 que forneça a tradução em L2 assim como *diferenciação de significado.* Esse

[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).

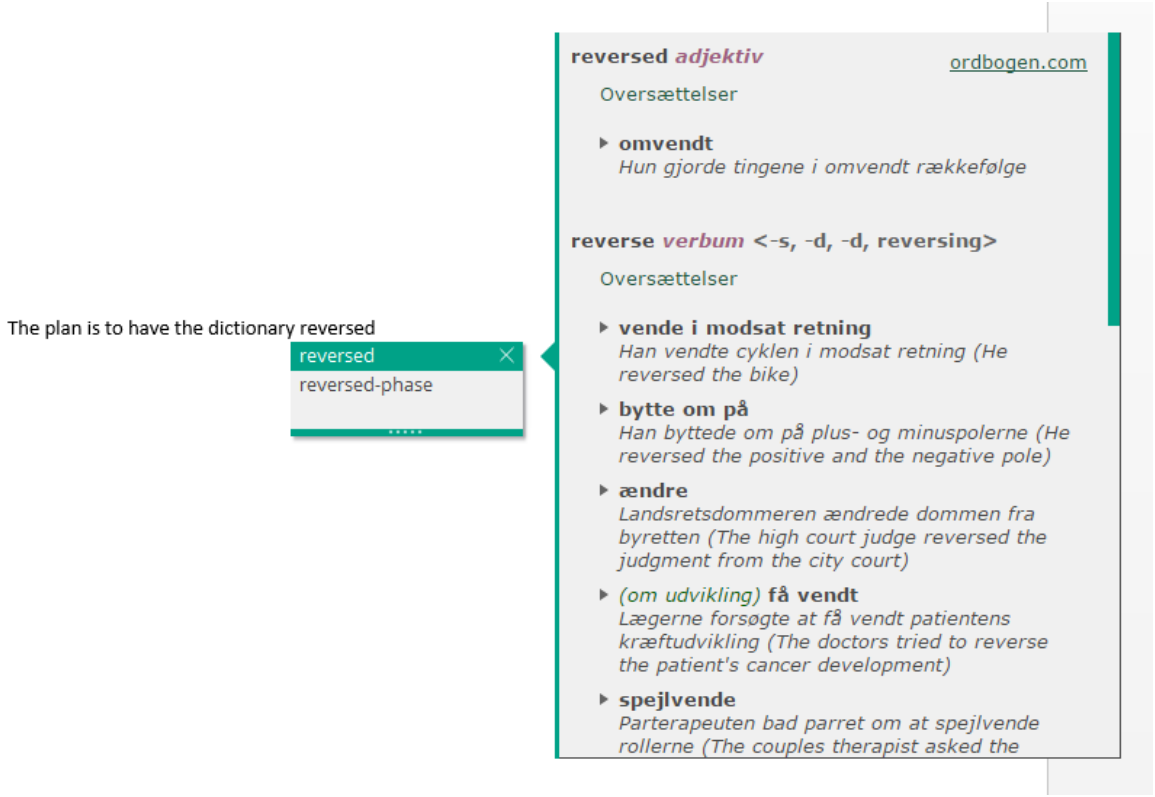
Figura 8 – A tupla "*the next to blande*" + equivalentes



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 *Kontekstеksempler*. 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.

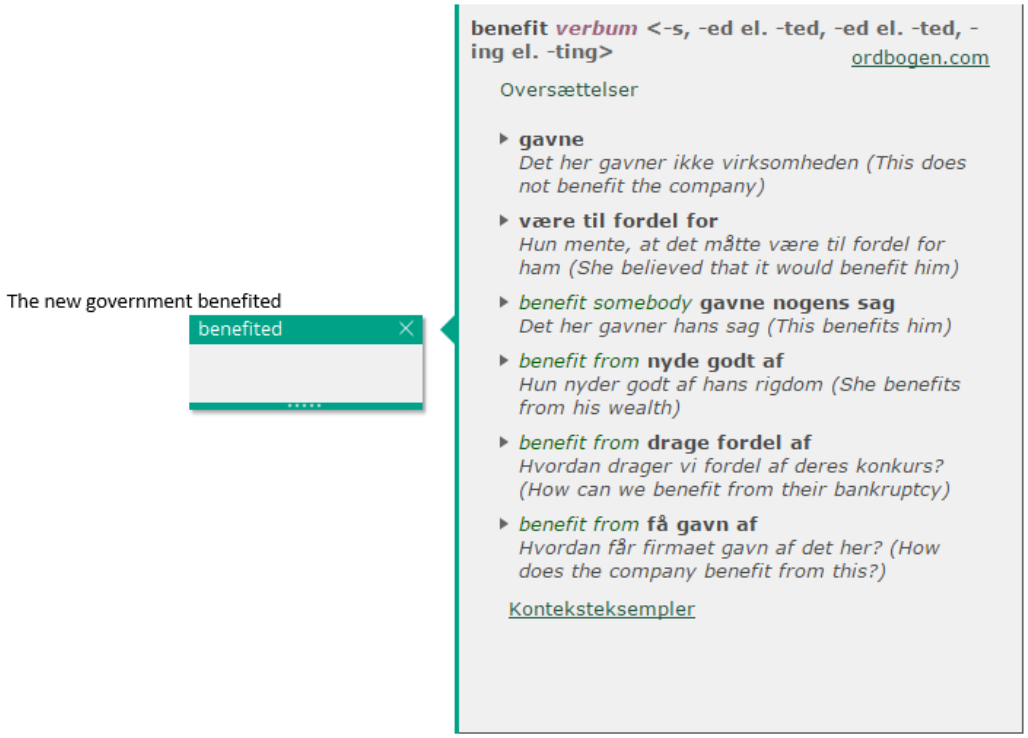
6) Quando o autor conhece uma palavra em L2, mas não tem certeza de como ela é escrita, ele precisará de informações sobre *ortografia* ou *autocorreção* ao digitar palavras em L2. Esse problema é completamente solucionado. Por um lado, a ferramenta oferece finalizações escritas corretamente depois que 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.

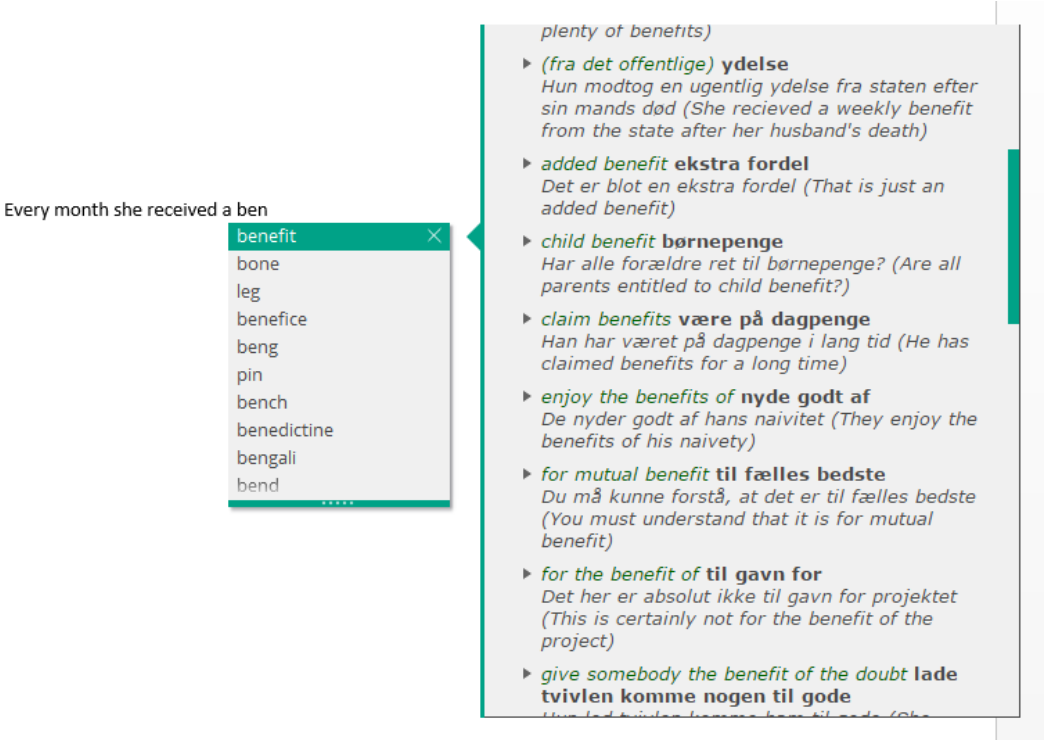
Figura 10 – A tupla "*The new government benefited*" + entrada do dicionário



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.

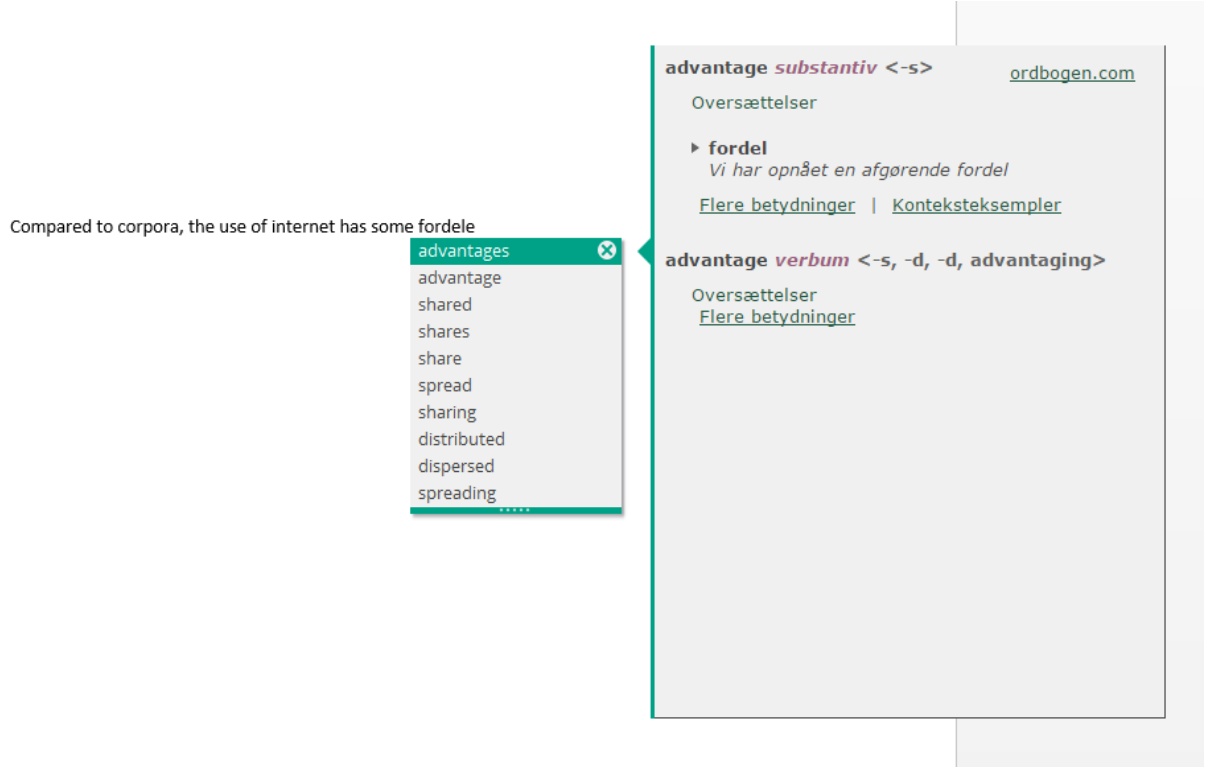
O uso de dicionários já existentes pode criar alguns problemas que afetam a qualidade geral da ferramenta. Hoje, a maioria dos dicionários ainda é concebida para fornecer assistência à produção de textos e à recepção de textos sem a necessária diferenciação. A concepção e a apresentação dos diferentes 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.

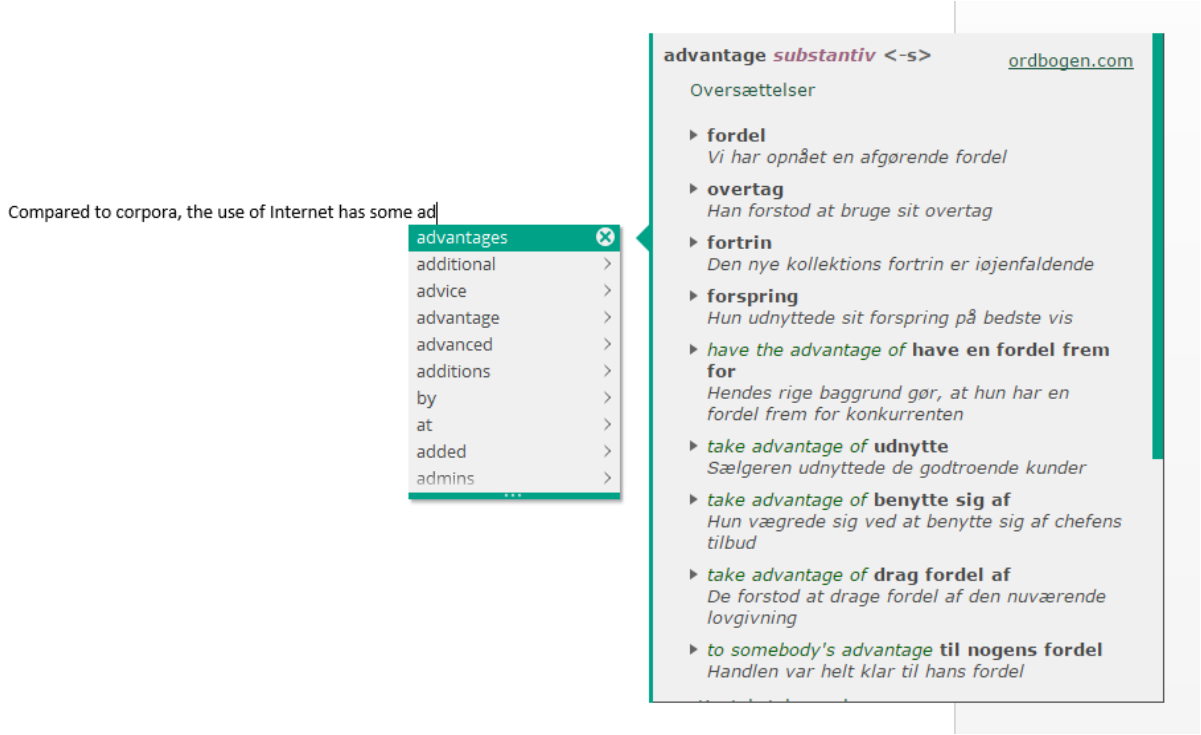
Na Seção 4.4, vimos como a mesma categoria de dados desempenhou o papel de ambos, explicação de significado e diferenciador de significado. Esse duplo papel pode criar alguns inconvenientes se não tiver sido previsto e levado em consideração no momento em que o conceito para os dicionários subjacentes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

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



Fonte: Acervo do autor

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



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.; FISKER, 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.; FISKER, 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.