

## UTILIZANDO O *SKETCH ENGINE* A FAVOR DO ENSINO-APRENDIZAGEM DE ESPANHOL COMO SEGUNDA LÍNGUA

*Amanda Maraschin Bruscato\**

**RESUMO:** A sociedade atual vivencia uma nova revolução ocasionada pelo avanço das Novas Tecnologias da Informação e Comunicação. Contudo, a prática docente não parece acompanhar as transformações das tecnologias (COLLINS & HALVERSON, 2018). Baseando-se na necessidade de mudança, este artigo tem como objetivo demonstrar como a tecnologia digital disponível para análise de *corpora* pode contribuir para o aprimoramento da prática do professor de línguas. Inicialmente, apresenta-se uma breve revisão sobre a linguística de *corpus* e a ferramenta de análise *Sketch Engine*. Após, explica-se um exemplo de proposta didática planejada para o ensino-aprendizagem de verbos utilizados na situação comunicativa de ida ao médico em espanhol. Em seguida, descreve-se a metodologia de pesquisa da linguística de *corpus* e analisam-se os dados obtidos no *Sketch Engine*. Foram utilizados quatro *corpora* para análise, sendo um geral e um especializado para a língua espanhola e outros dois para a língua portuguesa. Através do recurso *Word Sketch*, pesquisaram-se as colocações mais frequentes com o verbo *doler/doer* e com os nomes *dolor/dor*, *enfermedad/doença* e *enfermo/doente*. Discutem-se os problemas e soluções encontrados na pesquisa e conclui-se o artigo reafirmando a importante contribuição da linguística de *corpus* para o aprimoramento da prática docente.

**PALAVRAS-CHAVE:** Ensino de línguas; Língua espanhola; Linguística de corpus; *Sketch Engine*.

### Introdução

O surgimento das Novas Tecnologias de Informação e Comunicação (NTICs) como os computadores e a Internet tem transformado o mundo nas últimas décadas. De acordo com Collins e Halverson (2018), a sociedade presencia no século XXI uma revolução intitulada Revolução da Informação ou Revolução do Conhecimento com a mesma potência da Revolução Industrial.

---

\* Universidade do Algarve.

No entanto, os autores criticam que, apesar de as instituições educacionais criarem laboratórios de informática e adotarem as NTICs nas salas de aula, elas dificilmente repensam e reformulam as práticas básicas de ensino-aprendizagem tendo em vista as mudanças ocasionadas pela nova Revolução. Partindo desta crítica, este artigo pretende demonstrar como a tecnologia digital disponível para análise de *corpora* pode contribuir para que os professores de línguas aprimorem suas aulas.

Um *corpus* pode ser definido como um conjunto de textos autênticos em formato eletrônico selecionados de acordo com critérios externos para representar uma língua ou variedade linguística (MCENERY et al., 2006, p. 5). Ele é diferente de um arquivo de textos justamente por seu caráter representativo, ou seja, passível de generalização.

Em sua origem, a linguística de *corpus* era limitada pelas tecnologias disponíveis, dificultando que suas análises fossem de fato representativas da língua estudada. Ainda assim, manualmente foi compilado o *corpus Survey of English Usage* na metade do século XX com 1 milhão de palavras, que serviu posteriormente como base para a *Comprehensive Grammar of the English Language* (QUIRK et al., 1985).

O primeiro *corpus* eletrônico, também limitado devido ao espaço disponível para armazenamento de dados dos computadores da época, atingiu 1 milhão de palavras (Francis & Kucera, 1964). Ao final do século XIX, contudo, a linguística de *corpus* alcançou novas proporções com a compilação do *British National Corpus* (BURNARD, 1995), primeiro *corpus* a ter 100 milhões de palavras, e a escrita da *Longman Grammar of Spoken and Written English* (BIBER et al., 1999), gramática da língua inglesa baseada inteiramente na análise de um *corpus* de 40 milhões de palavras criado para tal fim.

Para ser representativo de determinada língua ou variedade linguística, é necessário que o *corpus* seja composto por textos autênticos produzidos por falantes nativos. Sinclair (1991) e Tognini-Bonelli (2001) são autores que defendem a necessidade da coleta de textos produzidos para autênticos propósitos comunicativos. Widdowson (1998), no entanto, sugere que o texto, assim que retirado de seu contexto e utilizado para outro fim – como o ensino –, perde sua autenticidade.

Independentemente de o texto perder ou não sua autenticidade original ao ser aproveitado para outro fim, é consenso que os textos utilizados para compilação de *corpora* devem ter um propósito original autêntico e ter sido produzidos por falantes nativos – a menos que seja um *corpus* de aprendizes. Um exemplo de *corpus* de aprendizes é o International Corpus of Learner English (GRANGER, 2003), com mais de 3 milhões de palavras escritas por estudantes de inglês de diversas universidades.

Há múltiplas possibilidades de trabalho com *corpora* para os professores de línguas. Eles podem compilar seus próprios *corpora* de aprendizes para analisar o que seus estudantes estão dizendo ou escrevendo na língua-alvo, assim como podem utilizar *corpora* gerais da língua que ensinam para descobrir o que de fato os nativos dizem ou escrevem. Independentemente do tipo de *corpus* que o docente decidir utilizar, é possível tomar duas atitudes em relação a ele, denominadas por Boulton (2010) como *hands-off* e *hands-on*. A primeira é utilizar os *corpora* para preparar materiais didáticos, e a segunda é levar os próprios alunos a pesquisarem nos *corpora*. Esta abordagem foi proposta e denominada por Johns (1991) como *Data-Driven Learning* (DDL), permitindo que os discentes sejam mais autônomos e atuantes no processo de ensino-aprendizagem.

Para este artigo, foi planejada uma atividade de pesquisa sobre verbos em espanhol voltada ao contexto de ida ao médico, que será explicada adiante. Caso os estudantes não tenham acesso às tecnologias digitais para realizarem a pesquisa proposta, há a alternativa de o docente levar as tabelas e exemplos aqui oferecidos para que os alunos as interpretem e discutam suas conclusões.

### Ferramenta de análise de *corpus*

A ferramenta de análise de *corpus* empregue na pesquisa realizada foi o *Sketch Engine* (KILGARRIFF et al., 2004). Ele é tanto um software como um serviço da web, dispondo de diversos *corpora* já anotados e possibilitando a criação de novos.

O *Sketch Engine* existe há 15 anos e é bastante utilizado na lexicografia, linguística computacional, análise do discurso, em pesquisas com tradução e ensino de línguas. Apesar

de hoje possuir diversas funções, sua origem está na função *Word Sketch*, criada em 2002 para auxiliar na elaboração de dicionários. *Word Sketch* é a síntese do comportamento gramatical e das colocações de determinada palavra (KILGARRIFF et al., 2014, p. 9). Ao selecionar a colocação, também é possível através da função Concordance visualizar o contexto em que cada item ocorre.

A ferramenta oferece um mês de serviço gratuito, sendo paga após este tempo. Outra alternativa, no entanto, é o *Sketch Engine* for Language Learning (SkELL), que oferece gratuitamente e sem necessidade de cadastro as colocações mais frequentes para a palavra procurada com 40 exemplos cada (BAISA; SUCHOMEL, 2014). Os resultados são limitados e a ferramenta está disponível apenas para o inglês, russo, alemão, italiano, tcheco e estoniano. Contudo, é gratuita, fácil de usar e logo estará disponível para outras línguas.

Caso a língua de ensino seja uma das disponíveis no SkELL, sugere-se que os professores realizem atividades nas quais os próprios alunos pesquisem em seus celulares ou computadores as colocações mais frequentes para determinadas palavras. Caso a língua seja outra e os estudantes não disponham de computadores e Internet, recomenda-se que o docente imprima os resultados encontrados por ele no *Sketch Engine* leve para interpretá-los e discuti-los com a turma.

Como há a possibilidade de utilizar o SkELL para a língua inglesa e já existe um livro didático voltado ao ensino de inglês com o *Sketch Engine* (THOMAS, 2014), serão apresentados aqui os resultados de uma pesquisa para o ensino-aprendizagem de verbos utilizados no contexto de ida ao médico em língua espanhola e comparados com os resultados em língua portuguesa. Antes, porém, de explicar a metodologia adotada para a pesquisa, serão definidos alguns termos utilizados pela linguística de *corpus* e necessários para analisar os dados oferecidos pela ferramenta.

As ferramentas de análise de *corpora* possuem como uma de suas funções a anotação automática dos *corpora*. A anotação diz respeito à adição de etiquetas gramaticais a todas as palavras do *corpus* de modo a permitir sua análise. Para esta pesquisa, será necessário saber que *V* é utilizado para verbo, *N* para nome, *A* para adjetivo e *S* para preposição, assim

como *subj* significa sujeito e *obj* significa objeto. Antes de a ferramenta computar a frequência das colocações, cada palavra tem seu lema identificado, ou seja, sua forma básica, encontrada nos dicionários. A ferramenta, então, contabiliza a quantidade de palavras, lemas e tokens, sendo estes cada palavra, número e sinal dentro do corpus.

Além destes termos, é importante distinguir um *corpus* geral de um *corpus* especializado. Um *corpus* geral de uma língua é bastante extenso e busca ser representativo de toda a língua ou variedade selecionada, coletando, portanto, textos dos mais diversos temas e gêneros textuais. Já um *corpus* especializado define um ou alguns domínios de estudo ou gêneros textuais e busca ser representativo apenas da especialidade escolhida. Para esta pesquisa, ambos serão utilizados.

### **Proposta didática**

A atividade sugerida neste artigo foi planejada para alunos do ensino médio que já possuam um conhecimento intermediário da língua espanhola. Tendo como base teórica o Interacionismo Sociodiscursivo proposto por Bronckart (1999) e seguido por Schneuwly e Dolz (2004) na proposta das sequências didáticas, a seguinte oficina foi construída a partir da situação comunicativa de ida ao médico.

Uma sequência didática é composta inicialmente pela apresentação do tema e por uma primeira produção textual pelos alunos. Após, são realizadas oficinas de modo a aprimorar a produção textual e avaliá-la ao fim da sequência.

Como sugestão de apresentação do tema, sugere-se que os alunos assistam ao vídeo *Ir al médico con tu mamá* e conversem sobre o que entenderam e sobre os momentos de tensão que ocorrem. Como produção textual, propõe-se que eles se dividam em trios e que cada grupo escreva um diálogo para, ao fim da sequência, gravar seu próprio vídeo sobre a situação de ida ao médico (seja de comédia, drama ou outro gênero).

Inicialmente, os estudantes deverão listar quais palavras acreditam serem típicas desta situação comunicativa. Após, deverão discutir e anotar quais colocações acreditam serem mais frequentes para cada palavra pensada.

Com base nesta atividade intuitiva, escreverão a primeira versão do diálogo. Após, será realizada a oficina com uso do *Sketch Engine*. Em uma tabela, os grupos deverão pesquisar no *Word Sketch* os lemas das palavras listadas anteriormente e anotar suas colocações mais frequentes, comprovando se sua intuição estava correta.

Cada grupo apresentará seus resultados para discutir com a turma e, após refletirem sobre as colocações mais frequentes, aprimorarão seu diálogo. Por fim, poderão gravá-lo e disponibilizá-lo na Internet.

Caso os estudantes não tenham acesso a computadores e Internet, o professor pode realizar a pesquisa e levar as tabelas para que os alunos analisem em aula. Pensando em já oferecer os resultados ao professor de língua espanhola, este artigo oferece a seguir os dados obtidos com a pesquisa.

## Metodologia

Este estudo utiliza a metodologia de pesquisa da linguística de *corpus* através do software *Sketch Engine*. Com o objetivo de demonstrar aos professores de línguas como a ferramenta de análise de *corpus* pode contribuir para suas aulas, foi realizada uma pesquisa com *corpora* das línguas portuguesa e espanhola para o estudo de verbos utilizados no contexto de ida ao médico.

Pensou-se como público-alvo aprendizes de espanhol cuja língua materna seja o português, mas as atividades podem facilmente ser revertidas para aprendizes de português cuja língua materna seja o espanhol. De modo a comparar os resultados de ambas as línguas, bem como comparar os dados obtidos com *corpus* especializado e *corpus* geral, foram utilizados quatro *corpora*.

Para a criação dos *corpora* especializados foi utilizada a função disponibilizada pelo *Sketch Engine* de criação de *corpus* com textos da web a partir de palavras-chave. Este método pode ser adotado por professores que trabalham com aulas direcionadas a temas específicos, caso desta pesquisa voltada à situação comunicativa de ida ao médico. De acordo com

o tema, as palavras-chave selecionadas para a língua espanhola foram *salud*, *enfermo*, *médico*, *hospital* e *cuerpo* e para a língua portuguesa foram *saúde*, *doente*, *médico*, *hospital* e *corpo*.

Os *corpora* gerais das línguas utilizados nesta pesquisa foram o Spanish Web 2018 (esTenTen18), criado a partir de textos espanhóis e hispânicos encontrados na web de fevereiro a abril de 2018, e o *Portuguese Web* 2011 (ptTenTen11), criado a partir de textos portugueses e brasileiros encontrados na web de março de 2011 a agosto de 2012. Enquanto o *corpus* do espanhol possui 20.306.642.991 tokens e 17.553.075.259 palavras, o *corpus* do português possui 4.622.750.491 tokens e 3.896.392.719 palavras.

O *corpus* especializado criado para o espanhol possui 246.537 tokens e 211.981 palavras, em torno de 0,001% do esTenTen18. Já o *corpus* especializado criado para o português possui 899.347 tokens e 706.878 palavras, quase 0,02% do ptTenTen11.

De acordo com Sardinha (2000, p. 346), um *corpus* é considerado pequeno se possui menos de 80 mil palavras, médio se possui de 250 a 1 milhão de palavras e grande se possui 10 milhões ou mais palavras. Segundo Sinclair (1991), para ser representativo de uma língua ou variedade linguística, o *corpus* deve ser o mais extenso possível, caso dos *corpora* gerais apresentados. Já os *corpora* criados com a ferramenta a partir de cinco palavras-chave são pequeno-médio (espanhol) e médio (português).

Para a análise, foi utilizada a função *Word Sketch* do *Sketch Engine*, que apresenta os seguintes métodos utilizados pela linguística de corpus: listagem das colocações léxico-gramaticais; computação de sua frequência; visualização de suas ocorrências; possibilidade de filtragem. Deste modo, as atividades realizadas foram: utilizar o *Word Sketch* para encontrar os verbos que possuem como complemento os lemas *enfermo/doente*, *enfermedad/doença*, *dolor/dor*, bem como os sujeitos de *doler/doer*; analisar os contextos das colocações encontradas através da função *Concordance*; comparar as frequências e contextos obtidos em uma tabela.

## Análise dos dados

Conforme explicado na seção anterior, para obter os dados nos quatro *corpora*, compará-los e analisá-los, foram escolhidos três nomes e um verbo relacionados ao contexto de ida ao médico. Os lemas pesquisados e os tipos de colocações buscadas para cada lema foram organizados no seguinte quadro:

**Quadro 1** – Esquematização da pesquisa

Lemas	Colocações
Doler / Doer	N suj de lema_V
Dolor / Dor	V obj lema_N
Enfermedad / Doença	V obj lema_N
Enfermo / Doente	V obj lema_N

**Fonte:** A autora

Quando se iniciou a pesquisa, contudo, percebeu-se que os lemas que aparecem como sujeitos de *doler* e *doer* também aparecem classificados como seus objetos, e que os lemas *enfermo* e *doente* apresentam resultados distintos dependendo de sua classificação como nome ou adjetivo. Sendo assim, os critérios para a pesquisa foram ampliados.

Ademais, para os *corpora* da língua portuguesa, o verbo *estar* não aparece junto aos outros verbos, mas à parte em uma categoria apenas com o verbo *ser*. Deste modo, para identificar as frequências de *estar doente*, foi necessário filtrar a categoria *N ser-estar doente*. Além desta, outra categoria que apareceu somente no *corpus* geral do português foi *verbo com se + doente\_N*, em que foi possível, através do recurso de filtragem, encontrar 54 ocorrências do lema *sentir* e as 252 ocorrências do lema *encontrar* no sentido buscado. Afinal, o significado de *encontrar* em *O poeta encontrava-se doente* é diferente de seu significado em *Apenas encontramos dois doentes*. Por fim, percebeu-se que as colocações de *padecer* e *sufrir/sofrer* seguidas pela preposição *de* estavam em uma categoria à parte das colocações dos mesmos verbos sem a preposição, também sendo necessário filtrá-la.



Após haver pesquisado as colocações frequentes nos quatro *corpora*, a frequência, ou seja, o número de ocorrências de cada colocação encontrada foi organizado no quadro a seguir:

**Quadro 2** – Colocações e frequências dos lemas pesquisados

Pesquisa	Resultado	Colocações	Língua Espanhola		Língua Portuguesa	
			Corpus Geral	Corpus Criado	Corpus Geral	Corpus Criado
Doler / Doer	Cabeza / Cabeça	Cabeza_N suj de doler_V / Cabeça_N suj de doer_V	449	-	127	-
		Doler_V obj cabeza_N / Doer_V obj cabeça_N	10.147	3	185	-
	Pecho / Peito	Pecho_N suj de doler_V / Peito_N suj de doer_V	336	-	68	-
		Doler_V obj pecho_N / Doer_V obj peito_N	2.288	-	62	-
	Rodilla / Joelho	Rodilla_N suj de doler_V / Joelho_N suj de doer_V	97	-	76	-
		Doler_V obj rodilla_N / Doer_V obj joelho_N	1.468	2	34	-
Dolor / Dor	Sentir / Sentir	Sentir_V obj dolor_N / Sentir_V obj dor_N	54.591	3	23.034	11
	Sufrir / Sofrer	Sufrir_V obj dolor_N / Sofrer_V obj dor_N	15.553	3	1.539	4

		Sufrir_V de_S dolor_N / Sofrer_V de_S dor_N	7.602	-	1.781	1
	Tener / Ter	Tener_V obj dolor_N / Ter_V obj dor_N	41.003	6	9.345	12
Enfermedad / Doença	Padecer / Padecer	Padecer_V obj enfermed- dad_N / Padecer_V obj doença_N	53.526	12	24	-
		Padecer_V de_S enfermed- dad_N / Padecer_V de_S doença_N	12.084	1	690	2
	Sufrir / Sofrer	Sufrir_V obj enfermed- dad_N / Sofrer_V obj doença_N	27.394	7	473	-
		Sufrir_V de_S enfermed- dad_N / Sofrer_V de_S doença_N	12.859	5	6.231	10
	Tener / Ter	Tener_V obj enfermed- dad_N / Ter_V obj do- ença_N	40.572	-	12.171	5
Enfermo / Doente	Caer / Cair	Caer_V obj enfermo_N / Cair_V obj doente_N	7.849	-	171	-
		Caer_V obj enfermo_A	2.785	-	-	-
	Encontrar / Encon- trar	Encontrar_V obj en- fermo_N / Encontrar_V com se + doente_N	3.496	-	252	-
		Encontrar_V obj en- fermo_A	3.095	1	-	-
	Estar / Estar	Estar_V obj enfermo_N / N estar doente_N	62.327	34	1.194	-
		Estar_V obj enfermo_A /	22.985	12	3.689	8

		N estar doente_A				
Poner / Ficar		Poner_V obj enfermo_N / Ficar_V obj doente_N	5.251	-	4.946	8
		Poner_V obj enfermo_A	3.116	-	-	-
Sentir / Sentir		Sentir_V obj enfermo_N / Sentir_V obj doente_N / Sentir_V com se + do- ente_N	2.740	3	243	-
		Sentir_V obj enfermo_A	1.920	-	-	-

Fonte: A autora

Percebe-se que os verbos e complementos encontrados no *corpus* geral da língua espanhola também costumam aparecer traduzidos no *corpus* geral da língua portuguesa. Ainda assim, provavelmente devido à diferença de tamanho entre os *corpora*, o *corpus* do espanhol possui bem mais ocorrências do que o do português, assim como os *corpora* gerais apresentam diversas colocações que não aparecem nos *corpora* especializados.

De modo a demonstrar como a linguística de *corpus* pode contribuir ao ensino-aprendizagem dos verbos em espanhol como segunda língua, também serão explicados os problemas que ocorreram na pesquisa e as ações efetuadas para solucioná-los. Pode-se dizer que a maior parte deles esteve relacionado a erros de classificação realizados pela ferramenta, os quais podem ser explorados para discussão em sala de aula.

Duas incongruências encontradas dizem respeito aos imprevistos referidos anteriormente: a classificação dos nomes que acompanham os verbos *doler* e *doer* ora como sujeitos e ora como objetos; e a anotação de enfermo e doente ora como nome e ora como adjetivo. O problema em si não é a diferença de classificação, mas os critérios adotados.

Para classificar o nome como sujeito ou complemento do verbo, a ferramenta parece basear-se simplesmente na ordem sintática dos elementos na frase, tendo como critério a posição do nome antes ou após o verbo. Desta forma, os nomes *pecho* e *peito* em *El pecho le dolía* e *O peito lhe doeu* são classificados como sujeitos, mas, nas frases *Me duele el pecho* e *Me dói o peito*, eles são classificados como objetos. Enquanto a classificação de sujeito ou objeto segue uma regra – ainda que equivocada –, a classificação de nome ou adjetivo para os lemas *enfermo* e *doente* não demonstra embasamento, uma vez que, em *Jorge Vergara está enfermo* e *Meu filho está doente*, os lemas são etiquetados como nomes, mas, em *Su hermano está enfermo* e *O povo está doente*, eles são etiquetados como adjetivos.

Em relação aos problemas mencionados, é possível extrair algumas conclusões. Uma delas é que a língua espanhola demonstra uma ordem sintática mais rígida para as construções com o verbo *doler* do que a língua portuguesa com o verbo *doer*. Em espanhol, a maioria das ocorrências do nome é posterior ao verbo, sendo, portanto, classificado equivocadamente pela ferramenta como objeto. Já em português, a preferência pela posição do nome em relação ao verbo não é tão evidente. Ainda assim, os dados são relevantes para o ensino-aprendizagem de espanhol, uma vez que impulsionam a discussão sobre como não é a ordem sintática que definirá se o nome adjacente ao verbo *doler/ doer* é seu sujeito ou objeto.

Sobre a anotação de *enfermo* e *doente* como nomes ou adjetivos complementares ao verbo, percebe-se que a mais comum é como nome, não havendo, em geral, a anotação como adjetivo para o português. Há, no entanto, uma exceção nos *corpora* da língua portuguesa: *doente* como complemento de *estar*, única colocação dentre as analisadas em que *doente* é etiquetado como adjetivo, inclusive apresentando mais ocorrências do que sua etiquetagem como nome em tal colocação. Novamente, a ferramenta demonstra um problema de anotação, visto que, nos casos encontrados, *enfermo* e *doente* tendem a ser adjetivos.

Outro equívoco da ferramenta está no momento de identificar o lema das palavras. Enquanto alguns equívocos são justificados por problemas ortográficos, como a escrita de *estuve* com a letra *b*, *estaba* com a letra *v*, *estábamos*, *están*, *estará* e *sentía* sem acento, impedindo

a identificação dos lemas *estar* e *sentir*, outros ocorrem devido à existência de uma mesma forma de palavra para dois lemas, caso dos verbos *sentir* e *sentar* em espanhol e em português. Desta forma, os lemas dos verbos em *Siento su dolor* e *Acho que sente dor* são erroneamente identificados como *sentar*.

Grande parte dos textos compilados para o *corpus* geral da língua espanhola foi retirada da *Wikipedia*, mas outros são da *Web*, apresentando por vezes abreviaturas, gírias e erros de ortografia. Nas frases *Estas minas estan re enfermas* e *¿Si estan enfermos, con q los curan?*, por exemplo, *están* está escrito sem acento, *re* é utilizado para intensificar *enfermas* e *que* é abreviado para *q*.

Apesar de a ferramenta apresentar alguns problemas como os mencionados acima, eles podem ser explorados em discussões em sala de aula. Pode-se, assim, analisar o uso da língua nas redes sociais, ensinar sobre como uma mesma forma de palavra pode servir para mais de um lema e como a ortografia correta das palavras é necessária para que as ferramentas automáticas possam anotar os *corpora* apropriadamente.

Os dois problemas de identificação dos lemas mencionados estão descritos numericamente no quadro a seguir:

**Quadro 3** – Falhas na identificação dos lemas

Colocação	Lema	Língua Espanhola		Língua Portuguesa	
		<i>Corpus</i> Geral	<i>Corpus</i> Criado	<i>Corpus</i> Geral	<i>Corpus</i> Criado
Estar_V obj enfermo_A / N estar doente_A	Estan	43	-	-	-
	Estar	22.919	12	3.689	8
	Estava	8	-	-	-
	Estube	15	-	-	-

Estar_V obj enfermo_N / N estar doente_N	Estabamos	16	-	-	-
	Estan	670	-	-	-
	Estar	61.632	34	1.194	-
	Estara	9	-	-	-
Sentir_V obj dolor_N / Sentir_V obj dor_N	Sentar	9.032	1	385	-
	Sentia	507	-	-	-
	Sentir	45.052	2	22.649	11
Sentir_V obj enfermo_A	Sentar	346	-	-	-
	Sentia	14	-	-	-
	Sentir	1.560	-	-	-
Sentir_V obj enfermo_N / Sentir_V obj doente_N / Sentir_V com se + doente_N	Sentar	200	-	16	-
	Sentia	12	-	-	-
	Sentir	2.528	3	227	-

Fonte: A autora

Os problemas com a ferramenta mencionados até o momento não estavam previstos no início da pesquisa. Ainda assim, acredita-se que eles possam contribuir para o ensino-aprendizagem de línguas, visto que impulsionam discussões e reflexões acerca de sua morfologia, sintaxe e semântica.

Quando se elaborou a atividade com os verbos utilizados no contexto de ida ao médico, houve dois principais objetivos. Para além de voltar o estudo linguístico a uma situação comunicativa que os alunos possam vivenciar, pensou-se em possíveis descobertas gramaticais necessárias aos aprendizes da língua.

A primeira descoberta, obviamente, é quais são as colocações mais frequentes nos *corpora*. Percebe-se, assim, que *sentir dolor/ sentir dor* é a colocação mais frequente para o lema

selecionado nos *corpora* gerais, enquanto *tener dolor/ter dor* é a mais frequente nos *corpora* especializados. Já para o lema *enfermedad/doença*, nos *corpora* do espanhol a colocação mais frequente é *padecer enfermedad*, enquanto no *corpus* geral do português é *ter doença* e no especializado é *sofrer de doença*. Em relação ao lema *enfermo/doente*, a colocação mais frequente em espanhol é *estar enfermo*, enquanto em português há uma frequência bastante próxima entre *estar doente* e *ficar doente*.

Indo além das colocações mais frequentes, houve outros objetivos gramaticais com esta atividade. Em relação ao verbo *doler/doer*, é importante que os estudantes percebam que o sujeito é a causa da dor e não a pessoa que a sofre. Esta necessidade não serve apenas a fim de conhecimento, mas para que a flexão do verbo faça sentido aos alunos. Assim, o verbo *doler*, assim como *doer*, é flexionado na terceira pessoa, sendo no singular no caso de *A mi burro le duele la cabeza* e no plural no caso de *A mi burro le duelen las rodillas*. A diferença deste verbo em português e espanhol, além da rigidez da ordem sintática mencionada anteriormente, é que, na língua espanhola, o pronome oblíquo átono deve acompanhar o verbo, enquanto, na língua portuguesa, ele é opcional.

Sobre os verbos *padecer* e *sufrir/sofrer* junto a *enfermedad/doença* e *dolor/dor*, percebe-se que, na língua espanhola, são mais frequentes as colocações sem a preposição *de*, mas o oposto ocorre na língua portuguesa. No entanto, as colocações *sofrer dor* e *sofrer de dor* em português não apresentam frequências tão distintas.

Com os resultados, os alunos podem discutir a diferença entre *sentir dolor/sentir dor* e *sentirse enfermo/sentir-se doente*. Através da análise dos *corpora*, também descobrem que os verbos *sentir*, *encontrar* e *poner* com *enfermo/doente* são pronominais, como *Uno de sus familiares se encuentra enfermo*, *Jiles se puso enfermo* e *Heracles se sintió enfermo* em espanhol e *A sua mulher encontra-se doente* e *Algumas crianças podem sentir-se doentes* em português.

Acerca dos verbos pronominais mencionados, é possível estudar sua diferença com o verbo *doler* – explicada por Abio (2008) –, cujo pronome não é reflexivo, mas objeto indireto. Além disso, pode-se estudar a posição dos pronomes em português e espanhol, bem como utilizar a opção de filtragem de *corpora* para separar os textos brasileiros dos

portugueses. Por fim, outra possibilidade de estudo com *corpora* é a irregularidade dos verbos em espanhol, que ocorre com *doler, sentir, encontrar, tener, poner, cair, padecer, sofrer e estar*.

Apesar de a análise de *corpus* contribuir para a análise gramatical, ela não precisa ter esse objetivo final. Conforme dito na introdução do artigo, o objetivo é utilizar as tecnologias digitais para reformular as práticas básicas de ensino-aprendizagem. Ao permitir que os alunos desenvolvam suas próprias conclusões a partir da análise de *corpus* e as discutam com os colegas, bem como incentivar o uso da ferramenta como auxiliar na produção de textos com reais propósitos comunicativos, a tecnologia contribui para uma aprendizagem mais significativa para o estudante, que passa a ser agente deste processo.

### Considerações finais

Este artigo pretendeu oferecer um exemplo de proposta didática em que a tecnologia digital disponível para análise de *corpora* pode contribuir para aprimorar as aulas de línguas. Conforme dito no decorrer do texto, a Revolução da Informação ou Revolução do Conhecimento que o mundo atualmente vivencia com o avanço das NTICs demanda a reformulação das práticas de ensino-aprendizagem. O professor, por conseguinte, passa a ser um facilitador, enquanto o aluno é o principal agente.

Foi apresentada como ferramenta para análise de *corpora* o *Sketch Engine*, ademais de sua versão reduzida e gratuita voltada ao ensino, o SKELL. Ambas versões possuem a função *Word Sketch* para pesquisa das colocações mais frequentes nos *corpora* gerais das línguas, mas viu-se que o SKELL é limitado e ainda não está disponível para a língua portuguesa ou a língua espanhola. Desta forma, foi planejada uma proposta didática do tipo *hands-on*, da abordagem DDL, voltada a aprendizes de espanhol cuja língua materna seja o português, e foram apresentados os resultados da pesquisa com quatro *corpora*, sendo um geral e um especializado para cada língua.

A partir da análise dos dados, conclui-se que os *corpora* especializados criados com a ferramenta são bem menos completos do que os *corpora* gerais e que, de fato, o tamanho



do *corpus* está relacionado com sua representatividade. Ainda assim, os *corpora* especializados oferecem uma amostra reduzida do funcionamento da língua e focam em temas específicos.

Foram sugeridos diversos tópicos de estudo a partir dos resultados obtidos. Ao analisar as colocações, é possível discutir a diferença entre forma de palavra e lema; regência e flexão verbais; e posição pronominal. No entanto, reforça-se que o estudo gramatical deve ser voltado a uma produção textual com real propósito comunicativo, como a sugerida na proposta didática.

A linguística de *corpus* pode auxiliar de múltiplas formas a prática docente, demonstrando como dados reais obtidos com falantes da língua são mais confiáveis do que a mera intuição e que devem, portanto, ser utilizados para o estudo linguístico. Espera-se, por fim, ter contribuído para a reflexão e a reformulação das práticas de ensino-aprendizagem de línguas.

### UTILIZANDO EL *SKETCH ENGINE* A FAVOR DE LA ENSEÑANZA-APRENDIZAJE DE ESPAÑOL COMO SEGUNDA LENGUA

**RESUMEN:** La sociedad actual experimenta una nueva revolución provocada por el avance de las Nuevas Tecnologías de la Información y Comunicación. Sin embargo, la práctica docente no parece seguir las transformaciones de las tecnologías (COLLINS & HALVERSON, 2018). Basado en la necesidad de cambio, este artículo tiene como objetivo demostrar cómo la tecnología digital disponible para el análisis de *corpora* puede contribuir a la mejora de la práctica de los profesores de lenguas. Inicialmente, se presenta una breve revisión de la lingüística de corpus y de la herramienta de análisis *Sketch Engine*. A continuación, se presenta un ejemplo de una propuesta didáctica planificada para la enseñanza-aprendizaje de los verbos utilizados en la situación comunicativa de ir al médico en español. Luego, se describe la metodología de investigación de la lingüística de *corpus* y se analizan los datos obtenidos en el *Sketch Engine*. Fueron utilizados cuatro *corpora* para el análisis, siendo uno general y uno especializado para la lengua española y dos para la lengua portuguesa. A través de la función *Word Sketch*, fueron pesquisadas las colocaciones más frecuentes con el verbo *doler/doer* y con los nombres *dolor/dor*, *enfermedad/doença* y *enfermo/doente*. Los problemas y las soluciones encontradas en la investigación son discutidas y por fin es reafirmada la importante contribución de la lingüística de *corpus* a la mejora de la práctica docente.

**PALABRAS CLAVE:** Enseñanza de Lenguas; Lengua Española; Lingüística de Corpus; *Sketch Engine*.

## REFERÊNCIAS

- ABIO, Gonzalo. Me llamo Pedro y Me duele la cabeza, no es lo mismo ni da igual: presentación de los verbos con pronombre en los libros de E/LE brasileños. In: CONGRESSO NORDESTINO DE ESPANHOL, 1., 2008, Recife. *Anais do I Congresso Nordeste de Espanhol*. Recife: Editora Universitária da UFPE, 2008. p. 165-177
- BAISA, V.; SUCHOMEL, V. SkELL – Web Interface for English Language Learning. In: WORKSHOP ON RECENT ADVANCES IN SLAVONIC NATURAL LANGUAGE, 8., 2014, Brno: *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Tribun EU, 2014. p. 63-70
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman grammar of spoken and written English*. London: Longman, 1999.
- BOULTON, Alex. Data-driven learning: Taking the computer out of the equation. *Language Learning*, Ann Arbor: University of Michigan, v. 60, n. 3, p. 534–572, 2010.
- BRONCKART, Jean-Paul. *Atividade de linguagem, textos e discursos: por um interacionismo sociodiscursivo*. São Paulo: EDUC, 1999.
- BURNARD, Lou (Ed.). *Users reference guide for the British National Corpus*. Oxford: Oxford University Computing Services, 1995.
- COLLINS, A.; HALVERSON, R. *Rethinking education in the age of technology: The digital revolution and schooling in America*. New York: Teachers College Press, 2018.
- FRANCIS, N.; KUCERA, H. *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence: Brown University, 1964.
- GRANGER, Sylviane. The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, Hoboken: Wiley-Blackwell, v. 37, n. 3, p. 538–546, 2003.
- JOHNS, Tim. Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, Birmingham: University of Birmingham, v. 4, 1991, p. 1-16, 1991.
- KILGARRIFF, A.; BAISA, V.; BUSTA, J.; JAKUBICEK, M.; KOVAR, V.; MICHELFETT, J.; RYCHLY, P. ; SUCHOMEL, V. *The Sketch Engine: ten years on. Lexicography*, Berlin: Springer, v. 1, n. 1, p. 7-36, 2014.
- KILGARRIFF, A.; RYCHLY, P. ; SMRZ, P. ; TUGWELL, D. (2004). *The Sketch Engine*. In: EURALEX, 11., 2004, LORIENT. *Proceedings of Euralex*. Lorient: Université de Bretagne Sud, 2004. p. 105-116

- MCENERY, T.; XIAO, R.; TONO, Y. *Corpus-based language studies: An advanced resource book*. London: Taylor & Francis, 2006.
- QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTVIK, J. *A comprehensive grammar of the English language*. London: Longman, 1985.
- SARDINHA, Tony. *Linguística de corpus: histórico e problemática*. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, São Paulo: Pontifícia Universidade Católica de São Paulo, v. 16, n. 2, p. 323-367, 2000.
- SCHNEUWLY, B.; DOLZ, J. *Gêneros orais e escritos na escola*. Campinas: Mercado de Letras, 2004.
- SINCLAIR, John. *Corpus, concordance and collocation*. Oxford: Oxford University Press, 1991.
- THOMAS, James. *Discovering English with the Sketch Engine*. Brno: Laptop Languages, 2014.
- TOGNINI-BONELLI, Elena. *Corpus linguistics at work*. Amsterdam: John Benjamins, 2001.
- WIDDOWSON, Henry. *Context, community and authentic language*. *TESOL Quarterly*, Hoboken: Wiley-Blackwell, v. 32, n. 4, p. 705–716, 1998.

Recebido em: 29/08/2019.

Aprovado em: 30/10/2019.