

Um léxico de verbos do dizer para tradutores – e considerações sobre a classificação dos verbos de elocução

A sense-based lexicon of reported-speech verbs – and remarks about semantic classification

Bianca Freitas Saburi Costa¹
PUC-Rio
bianca.saburi@gmail.com

Cláudia Freitas²
PUC-Rio
claudiafreitas@puc-rio.br

Resumo: Profissionais que traduzem textos do inglês para o português frequentemente se deparam com uma exigência das editoras quanto à tradução de diálogos: diversificar os verbos de elocução, com o intuito de evitar a repetição do verbo “dizer”. Em resposta, apresentamos o DISSE, um léxico de verbos de elocução criado a partir de grandes *corpora*, e organizado quanto ao sentido. Neste artigo, detalhamos o desafio de distribuir 293 verbos em grupos de sentido, tendo em vista facilitar sua consulta. Para tanto, realizamos dois testes, utilizando uma ferramenta desenhada para tirar proveito da anotação linguística como instrumento para o estudo de uma língua. Todos os resultados estão públicos para a consulta.

Palavras-chave: Verbos de elocução; Tradução; Classificação semântica.

Abstract: Professionals who translate texts from English to Portuguese often come across a specific demand from their clients regarding the translation of dialogues: they are asked to vary the reporting verbs, in order to avoid repeating the verb “dizer” (say). As a response to such demand, we put forward DISSE, a reporting verbs glossary created from large *corpora*. In this article, we discuss in detail the challen-

¹ Mestre em Estudos da Linguagem – PUC-Rio.

² Professora do Programa de Pós-Graduação em Estudos da Linguagem/PUC-Rio.

ge of distributing 293 verbs into distinct groups of meaning, in an attempt to make consultation easier. To establish those groups, we carried out two tests, using a tool that was designed to make use of linguistic annotation as an instrument for studying a language. All results are publicly available.

Keywords: Reported speech verbs; Translation; Semantic classification.

Introdução

Profissionais que traduzem textos do inglês para o português, principalmente nos âmbitos literário e jornalístico, deparam-se com uma exigência, por parte de editoras e agências de tradução, no que se refere à tradução de diálogo: evitar a repetição dos verbos que introduzem a fala de personagens. Em inglês, é extremamente comum que a sinalização de discurso relatado seja feita pelo verbo *say* em sua forma pretérita, *said*. Ao traduzir para o português, os tradutores são orientados a diversificar os verbos introdutores de discurso, empregando outros verbos de elocução. Mas, afinal, no português, temos tanta variedade de verbos para introduzir relato? Que verbos podem desempenhar essa função? Perguntas como essas levaram ao DISSE, um léxico de verbos de elocução elaborado com o intuito de facilitar a tarefa tradutória no que concerne à tradução de diálogos. Classificado quanto a classes de sentido, esperamos que o léxico sirva como auxílio ao tradutor, que, por conta da orientação de editoras e agências, deve lidar com uma variação lexical que não existe no texto fonte.

A elaboração do DISSE é o resultado de dois procedimentos: (a) a obtenção de uma lista de verbos de elocução efetivamente usados em português, que totaliza 293 verbos; (b) a organização dessa lista em agrupamentos de sentido que facilitem a consulta. O primeiro processo está detalhadamente descrito em Saburi Costa e Freitas (2017), e nos concentramos aqui no segundo procedimento, uma vez que a apresentação de uma lista tão grande é de utilidade limitada para um tradutor se não estiver organizada. A proposta final de classificação passou por dois testes, durante os quais encontramos diversos desafios.

Para nos guiar ao longo do processo, adotamos o ângulo sugerido pelos escritos tardios de Wittgenstein, expresso sobretudo nas *Investigações Filosóficas* (1953), e que comparece também, no contexto lexicográfico-computacional, no trabalho de Kilgariff (1997). Acreditamos, assim, que o sentido não é uma propriedade intrínseca das palavras, mas uma abstração, que só irá se concretizar no uso – traduzido como ocorrências da palavra em um *corpus* – e enquanto decorrência de algum objetivo ou tarefa³. Nessa perspectiva, sentido é materializado⁴ por lexicógrafos - a partir da observação e agrupamento das linhas de concordância em um corpus -segundo um objetivo específico, que é o de fazer dicionários. E justamente porque diferentes objetivos podem levar a diferentes sentidos, destacamos que o objetivo de nossa classificação é funcionar como um apoio ao tradutor.

A tarefa de classificação dos verbos, cerne deste trabalho, se aproxima da anotação linguística, prática comum na linguística computacional e que consiste na atribuição de uma categoria a um segmento

³ Veja-se Martins (2000) para uma discussão aprofundada sobre a questão do sentido na perspectiva wittgensteiniana.

⁴ A partir da observação e agrupamento das linhas de concordância em um corpus.

de texto, em contexto. Por isso, trazemos também algumas reflexões sobre a tarefa de anotação semântica que nos ajudam a entender o tamanho do desafio que enfrentamos e a justificar as decisões tomadas.

O presente artigo propõe reflexões acerca do processo de classificação semântica de verbos do dizer, motivadas pela elaboração de um léxico. Para tanto, discorremos brevemente sobre o processo de compilação dos verbos, para logo em seguida apresentarmos algumas propostas atuais de classificação semântica, atrelando à discussão considerações sobre procedimentos de anotação linguística. Apresentamos então nossa proposta de classificação, em duas etapas, entremeada com reflexões à atividade de categorização. Também exploramos possíveis desdobramentos do DISSE no contexto da linguística computacional e das humanidades digitais e, por fim, apresentamos nossas considerações finais e possibilidades futuras a partir do DISSE.

Compilação dos verbos do DISSE

Para a elaboração do DISSE, o primeiro passo foi uma busca exaustiva por verbos de elocução. Após uma consulta inicial a gramáticas e guias de estilo, decidimos basear nosso trabalho em grandes *corpora*, fazendo uso do vasto material disponível para a língua portuguesa. Para dar início ao processo, (a) utilizamos o COMPARA (Frankenberg-Garcia e Santos, 2002), um *corpus* paralelo bilíngue e bidirecional português-inglês, que contém trechos de obras literárias publicadas por editoras; e, em um segundo momento, (b) utilizamos *corpora* monolíngues disponibilizados para consulta no serviço AC/DC (Costa *et al.*, 2009), mantido pela Linguatca (Santos, 2011). Assim, nossa compilação partiu de uma lista de verbos de elocução já utilizados por tradutores profissionais em traduções literárias, em que se buscou, no COMPARA, por traduções possíveis para *said* (no original) em contextos de diálogo em livros publicados no Brasil e em Portugal.

A análise de mil linhas de concordância forneceu 59 verbos diferentes, como ilustra o quadro 1, que contém, ao lado do lema de cada verbo, a sua frequência de utilização. Destes 59, escolhemos 6 verbos: três verbos de elocução prototípicos (*dizer*, *perguntar* e *responder*); e três reconhecidamente polissêmicos (*admitir*, *contar* e *continuar*) para atuarem como sementes (ou disparadores) na busca, no *corpus* CHAVE (Santos e Rocha, 2005), por padrões léxico-sintáticos típicos dos verbos de elocução. Essa etapa nos forneceu 10 padrões. A aplicação dos padrões no *corpus* Floresta (Freitas *et al.*, 2008), por sua vez, trouxe uma lista de quase 600 verbos diferentes. Descartamos verbos com apenas uma ocorrência e, após uma análise manual dos resultados, com verificação das linhas de concordância a fim de confirmar que as ocorrências, de fato, se referiam a situações de elocução, obtivemos uma lista com 293 verbos. Em Costa e Freitas (2017) relatamos detalhadamente todo o processo de obtenção dos verbos.

Classificações semânticas de verbos

Uma lista de quase 300 verbos, ainda que útil, é de difícil manejo. A opção pela organização dos verbos em ordem alfabética é pouco funcional no contexto da prática tradutória, uma vez que a orde-

Quadro 1. Lista de traduções para a forma *said* em 958 ocorrências, compilada no COMPARA

59 TRADUÇÕES DE SAID									
dizer	561	declarar	7	falar	3	querer saber	2	escrever	1
responder	88	insistir	6	informar	3	admirar-se	1	espantar-se	1
perguntar	60	interromper	6	observar	3	admitir	1	indagar	1
omissão	53	prosseguir	6	retrucar	3	advertir	1	justificar-se	1
comentar	20	repetir	5	ripostar	3	anuir	1	lembrar	1
explicar	17	pedir	4	agradecer	2	anunciar	1	prometer	1
acrescentar	9	replicar	4	argumentar	2	balbuciar	1	saudar	1
afirmar	9	assegurar	3	contar	2	berrar	1	sondar	1
contrapor	9	comunicar	3	cumprimentar	2	brindar	1	sublinhar	1
exclamar	9	concluir	3	esclarecer	2	censurar	1	sugerir	1
retorquir	9	confessar	3	propor	2	concordar	1	tornar	1
continuar	8	confirmar	3	queixar-se	2	desabafar	1		

nação alfabética pressupõe que o consulente já sabe o item que procura (perspectiva semasiológica), e o que buscamos propor consiste justamente na situação oposta: encontrar um verbo que não se sabe exatamente qual é (perspectiva onomasiológica). A partir da ocorrência de *said* no original, no contexto em que aparece, o tradutor tem em mente uma noção aproximada de sentido. Por isso, acreditamos que a organização da lista em grupos de sentido seria a mais proveitosa para os tradutores.

Tentativas de classificação de verbos de elocução não são novidade. Na internet e até mesmo em alguns manuais de redação, é possível encontrar tais listas, porém sem qualquer organização a não ser a ordem alfabética. A exceção é o manual de Garcia (2010), que propõe uma classificação semântica para esses verbos, incluindo nove categorias. O autor reconhece, entretanto, que essas poucas categorias não dão conta dos muitos verbos que podem ser empregados como *dicendi*. Garcia explica que as categorias compreendem apenas os verbos *dicendi* mais comuns, porém ressalta que há muitos outros, que são “mais específicos, mais caracterizadores da fala” (Garcia, 2010, p. 149) e que não se encaixam nas categorias por ele sugeridas, como o verbo *sussurrar*, o que incita a necessidade de se criar novas categorias, de modo a abranger não só os verbos tradicionalmente usados como introdutórios de discurso direto ou indireto, mas também aqueles que são ocasionalmente empregados dessa maneira.

De um ponto de vista linguístico, o trabalho de Levin (1993) apresenta uma classificação de verbos que nos poderia auxiliar. A autora apresenta classes que são “sintaticamente relevantes” e “semanticamente coerentes” (Levin, 1993, p. 22). Dentre as 49 classes de verbos propostas, está “verbs of communication”, bastante ampla e, dentro dessa, a subclasse “say verbs”, que inclui diversos verbos de nossa lista. Porém, os critérios que demarcam cada uma das subcategorias não são sempre claros, e outras subcategorias parecem abrigar alguns dos verbos que consideramos (devido ao uso atesta-

do) verbos de elocução, como os “verbos de reclamar” e os “verbos de aconselhar”. Ainda em Levin (1993), para cada classe, uma lista de verbos é fornecida como exemplo, mas tais listas não têm a pretensão de exaustividade, como frisa. Em Levin, o objetivo é distribuir os verbos em classes semânticas segundo o seu comportamento sintático – e a opção de privilegiar a sintaxe se deve, segundo a autora, à dificuldade de identificação de significados possíveis com base apenas na intuição ou na definição de dicionários. Aqui, nosso objetivo, mais modesto, não é propor uma classificação geral dos verbos, mas distribuir os verbos de elocução em grupos de sentido que sejam relevantes para tradutores^{5,6}.

Wordnets também podem ser lembradas como recursos lexicais em que o critério semântico é elemento crucial da estruturação. Uma wordnet agrupa os diferentes sentidos de uma palavra em um *synset*, e esse agrupamento é um trabalho lexicográfico essencialmente humano. No entanto, se boa parte dos verbos que encontramos não constam de dicionários com o sentido de verbo de elocução⁷, não há razão para esperar que seja diferente nas wordnets (para uma apreciação e comparação das wordnets dedicadas à língua portuguesa, veja-se Oliveira *et al.*, 2015). Nossa tarefa envolve a classificação dos verbos a partir do contexto em que estão inseridos – e é apenas ele (o contexto) que nos lembra que o verbo em questão pode ser (e, de fato, foi) usado como verbo de elocução. Por isso, de certa maneira, nosso trabalho se alinha à atividade de anotação linguística, especificamente, de anotação semântica.

A atividade de classificar um segmento do texto em contexto está no cerne da tarefa de anotação linguística (Garside *et al.*, 1997). No entanto, enquanto a anotação está atrelada a um esquema de anotação e a um *corpus*, aqui nossa “anotação” está a serviço da organização de uma lista de verbos – o que não significa que a classificação que propomos não possa, futuramente, ser aplicada como um esquema de anotação.

Em um projeto de anotação, a maneira pela qual se verifica de que maneira as classes/etiquetas estão sendo atribuídas aos segmentos de texto é a concordância interanotadores. A avaliação da concordância consiste na comparação entre as diferentes classificações feitas pelos anotadores (humanos) ou entre eles e um gabarito⁸. Porque verifica o grau de consistência com relação a uma dada classificação, o valor da concordância interanotadores pode funcionar também como uma maneira indireta de avaliar o nível de dificuldade de uma determinada tarefa, e os valores das concordâncias são invariavelmente mais baixos quando a anotação envolve a atribuição de sentidos e, mais baixos ainda, quando a anotação de sentidos envolve verbos, justamente devido à alta polissemia (ver Artstein e Poesio (2008) para uma ampla discussão sobre índices de concordância).

Para ilustrar os desafios da anotação semântica, retomamos brevemente a experiência do SemCor (Landes *et al.*, 1998; Baker *et al.*, 2017) um subconjunto do *Brown Corpus* anotado semanticamente segundo os *synsets* da WordNet de Princeton (Fellbaum, 1998). O trabalho foi feito tomando por base

⁵ Ainda que sua utilidade não se restrinja aos tradutores, o recorte é fundamental para nortear certas decisões de classificação.

⁶ Para um maior detalhamento da proposta de Levin com relação aos verbos de elocução, veja-se Freitas (2016).

⁷ Veja-se, por exemplo, o verbo *brincar*: “‘Não me lembro deste jogo’, brincou McEnroe.”

⁸ De maneira mais precisa, a concordância interanotadores foi idealizada como uma maneira de se avaliar a confiança – que podemos traduzir por “consistência” – na anotação manual de um corpus. A confiança materializa a validade de um esquema de anotação, isto é, a validade de uma maneira de analisar e classificar os dados linguísticos. No entanto, consistência não significa correção, ou seja, é possível ser consistente no erro, o que não é capturado pelos índices de concordância.

a hipótese – agora vista como ingênua, nas palavras dos autores⁹ – de que a anotação semântica seria um trabalho simples de alinhamento, limitando-se a uma inversão do trabalho de lexicógrafos, que usam a intuição para criar entradas lexicais a partir da inspeção de linhas de concordância. No entanto, continuam Baker *et al.* (2017), a experiência mostrou que o trabalho não tinha qualquer aspecto de simplicidade: os anotadores relataram que, em vários casos, nenhum dos sentidos disponíveis para o alinhamento parecia se encaixar, ou que o sentido da palavra em questão era capturado por mais de um sentido da WordNet. Os autores sugerem – e concordamos – que frequentemente os sentidos das palavras resistem a uma representação discreta e estável como a dos dicionários. Como observa Wittgenstein (1963, §71), uma imagem pouco nítida “é justamente aquela de que, com frequência, precisamos”, e como comenta Stock (1983 *apud* Kilgarriff, 1997, p. 101), “it is precisely the lack of clarity in our use of the word culture which makes it such a handy word to have at one’s disposal.”¹⁰

Por fim, para além da dimensão empírica que a anotação linguística oferece para o objetivo pretendido – em nosso caso, ilustrando a sua dificuldade – lembramos que pode haver uma dimensão teórica subjacente, quando a anotação funciona como um teste de hipóteses relativo às categorias propostas para abordar uma questão, e esse aspecto também nos interessa para lidar com a atividade de classificação semântica:

A anotação se apresenta [...] como um teste empírico desenhado à maneira clássica: criam-se hipóteses (categorias provisórias), e verificam-se as hipóteses (as categorias são aplicadas no *corpus* – o processo de anotação acontece). Durante a anotação/verificação, as hipóteses podem ser confirmadas (a anotação ocorre exatamente como previsto) ou, por outro lado, os dados podem levar à reformulação das categorias iniciais. As categorias (hipóteses) recém-criadas são novamente verificadas no *corpus*, e o processo recomeça. (Freitas, 2015, p. 49)

A dificuldade experimentada por Garcia (2010), Levin (1993) e Landes *et al.* (1998), e materializada pelos baixos valores de concordância, já havia sido sentida, décadas antes, por John Austin (1962), quando se debruça sobre os verbos de elocução no contexto dos Atos de Fala. Uma vez que a tentativa de classificação dos verbos proposta por ele – e o seu reconhecido “fracasso” – provocam reflexões relevantes para o desenvolvimento do DISSE, apresentaremos uma análise mais detalhada desse trabalho após a nossa primeira tentativa de classificação.

Primeira tentativa de classificação dos verbos de elocução

Tirando proveito da experiência profissional com tradução de uma das autoras deste artigo, propomos inicialmente 21 classes para organizar os 59 verbos obtidos com o COMPARA – isto é, apenas

⁹ “The work was based on the assumption—now clearly recognized as having been naive—that semantic annotation would be a simple and straightforward task.” (Baker *et al.*, 2017, p. 702)

¹⁰ “It is precisely the lack of clarity in our use of the word culture which makes it such a handy word to have at one’s disposal. It offers, as it were, semantic extras just because in most uses its possible meanings are not clearly disambiguated. [...] What can the dictionary maker do to reflect this state of affairs? [...] They do not, cannot by their very structure, show that there is slippage between some of the senses that they give but not between others.” (Stock, 1983 *apud* Kilgarriff, 1997, p. 101)

aqueles verbos que já haviam sido traduzidos como alternativas ao *said*. O quadro 2 apresenta as classes criadas e os verbos incluídos em cada classe.

Quadro 2. Classificação inicial dos 59 verbos de dizer obtidos no COMPARA, distribuídos em 21 grupos de sentido

CLASSE	VERBOS						
acordo	anuir	concordar					
acréscimo	acrescentar	continuar	prosseguir				
afirmação	afirmar	comentar	contar	declarar	dizer	falar	
censura	censurar						
confirmação	assegurar	confirmar	repetir				
confissão	admitir	confessar	declarar	desabafar			
contestação	argumentar	contrapor	insistir	queixar-se	replicar	retorquir	retrucar
destaque	insistir	observar	sublinhar				
exclamação	berrar	exclamar					
explicação	argumentar	concluir	esclarecer	explicar	justificar-se	observar	
hesitação	balbuciar						
informação	anunciar	comunicar	contar	declarar	informar		
interrupção	interromper						
lembrança	lembrar						
opinião	concluir	insistir					
pedido	pedir						
pergunta	indagar	perguntar	querer saber	sondar			
resposta	replicar	responder	retorquir	retrucar	ripostar		
sugestão	advertir	propor	sugerir				
surpresa	admirar-se	espantar-se					
outros	agradecer	brindar	cumprimentar	escrever	prometer	saudar	

Para validar a classificação, desenvolvemos o seguinte teste¹¹: apresentamos 59 linhas de concordância, uma para cada verbo, e os respondentes (tradutores profissionais e amadores) deveriam classificar o verbo em destaque conforme nossas classes, sempre levando em conta o contexto. Os respondentes poderiam, ainda, sugerir novas classificações, caso acreditassem que nenhuma das classes propostas desse conta do verbo em questão. A Figura 1 apresenta um trecho do teste.

Dez pessoas participaram da atividade. Os resultados puseram em xeque a classificação proposta¹², e poucas classes foram validadas. Estabelecemos como critério para validação a concordância de 80% dos respondentes. Em outras palavras, se oito ou mais indivíduos atribuísem a mesma classe a um determinado verbo no contexto da frase, essa classe seria, portanto, validada (mesmo que não

¹¹ De acordo com a perspectiva teórica assumida no trabalho, não há como acreditar que qualquer classificação bem-sucedida seja diferente de “consensual”. Como consenso é fruto de concordância, a consulta a especialistas em linguagem, sobretudo tradutores, objetiva justamente verificar o quanto a classificação proposta é compartilhada. Além disso, o método proposto se aproxima das tarefas de anotação, já mencionadas, em que substituímos o *corpus* por uma frase.

¹² Em Freitas (2016) estão disponíveis as respostas dadas por todos os participantes.

tivesse sido sugerida por nós). Trata-se de um valor alto para a concordância, se levarmos em conta as já mencionadas dificuldades da anotação semântica de verbos. Dos 59 verbos, apenas 18 tiveram suas classes confirmadas, e os listamos no quadro 3.

Figura 1. Trecho da primeira versão do teste

TRADUÇÃO	CATEGORIA(S)	SUGESTÕES
– Obrigada, mas acho que não – respondeu com dificuldade.		
– E se eu não puder? – perguntei .		
– Eu não sei o que quero – confessou Alistair com ar de desânimo. – Só não quero que tudo isto continue.		

Fonte: Freitas (2016).

Quadro 3. Lista de verbos e os grupos de sentido validados pelo primeiro teste

Verbo	Classe	Concordância
pedir	PEDIDO	90%
sublinhar	DESTAQUE	90%
acrescentar	ACRÉSCIMO	90%
anuir	ACORDO	80%
replicar	CONTESTAÇÃO	80%
censurar	CENSURA	90%
perguntar	PERGUNTA	80%
confessar	CONFISSÃO	90%
propor	SUGESTÃO	90%
querer saber	PERGUNTA	90%
balbuciar	HESITAÇÃO	80%
confirmar	CONFIRMAÇÃO	80%
sugerir	SUGESTÃO	100%
admirar-se	SURPRESA	80%
exclamar	EXCLAMAÇÃO	90%
declarar	AFIRMAÇÃO	80%
indagar	PERGUNTA	80%
lembrar	LEMBRANÇA	80%

Apenas um verbo teve sua classe validada por todos os participantes, o verbo *sugerir*. E, mesmo nesse caso, as pessoas não deixaram de atribuir, simultaneamente, outras classes preestabelecidas, ou, ainda, de sugerir novas classificações. Assim, foram contabilizadas, no total, sem repetições, 304 classes (havia, originalmente, 51), o que corresponde a uma média de 5,24 classes distintas para o mesmo verbo.

Os resultados obtidos refletiram as possibilidades da classificação múltipla: *indagar* e *lembrar*, verbos aos quais foram atribuídos o menor número de classes, foram enquadrados em duas classes cada. Por outro lado, *agradecer*, *prosseguir* e *queixar-se* encaixaram-se, segundo os participantes, em nove classes cada um. O Quadro 4 mostra os grupos atribuídos a cada um desses cinco verbos.

Quadro 4. Lista de classes atribuídas aos verbos que menos e mais receberam classificações diferentes

OCORRÊNCIA	CLASSES ATRIBUÍDAS
– Porquê? – indagou Mr. Walsh.	CONTESTAÇÃO, PERGUNTA
– Até os ovos – lembrou alguém.	ACRÉSCIMO, LEMBRANÇA
– Foi muito simpático da sua parte – agradeceu Bernard.	AFIRMAÇÃO, AGRADECIMENTO, CONFIRMAÇÃO, CORTEJO, DESTAQUE, EXCLAMAÇÃO, GRATIDÃO, OPINIÃO, RESPOSTA
– No entanto gosto do azul – prosseguiu , a cliente – e o verde é muito bonito.	ACORDO, ACRÉSCIMO, AFIRMAÇÃO, CONTRAPOSIÇÃO, INFORMAÇÃO, INTERRUPTÃO, NEUTRALIDADE, OPINIÃO, SUGESTÃO
– De qualquer modo, é muito frio aquilo em Sea Cliff, em toda aquela ponta do Noroeste de S. Francisco – queixou-se .	CENSURA, CONCLUSÃO, CONFISSÃO, EXCLAMAÇÃO, LAMENTO, OPINIÃO, QUEIXA, RECLAMAÇÃO, RESPOSTA

Apesar de termos antecipado a possibilidade de os verbos se enquadrarem em mais de um grupo, consideramos o valor médio de grupos por verbo um tanto alto, o que sem dúvida dificultaria a consulta do tradutor ao léxico. Com muitos verbos repetidos em grupos de sentido diferentes, a classificação pode perder o seu propósito, já que haveria várias listas longas com os mesmos verbos. Além do número relativamente baixo de validações e da média alta de grupos por verbo, ressaltamos também o fato de que nenhum verbo recebeu apenas uma classe.

Com os resultados desse primeiro teste, ficou evidente que uma classificação granular demais (21 classes para 59 verbos) suscita mais discordâncias e não facilita a tarefa do tradutor, pelo contrário: ele se vê obrigado a decidir entre diferentes nuances de sentido, sendo levado a dúvidas que talvez não existissem. Ao se deparar com o verbo *said*, o tradutor seria obrigado a escolher dentre classificações refinadas, em geral desnecessárias para encontrar uma tradução que se encaixasse no contexto em questão. Ficamos tentados a pensar que um número maior de categorias torna um sistema de classificação mais sensível às diferentes situações. No entanto, à medida que o número de categorias se aproxima do número de situações elencadas, toda a ideia de categorização se esvai, perdendo sua utilidade na prática e podendo criar um obstáculo a mais à consulta. Ficou claro para nós, então, que uma reformulação da classificação deveria, necessariamente, envolver grupos mais abrangentes, ou seja, mais genéricos.

Considerações sobre a atividade de classificação

Para lidar com os desafios da classificação, nos apoiamos em pensadores como Nietzsche (1873) e Wittgenstein (1953), que nos lembram que classificar envolve agrupar aquilo que *não é* semelhante. A classificação, como detalha Ellis (1993), é um processo ativo de construção de homogeneidade, uma construção *interessada*, que acontece por meio do descarte de *certas* diferenças – apenas aquelas que nos interessa descartar. Em outras palavras: a atividade de classificação envolve simplificar o que é complexo, para isso descartando algumas características, e igualando o que não é igual. Esse processo não é natural: o que será descartado, por um lado, e a homogeneidade construída, por outro, são sempre decorrências do interesse de quem classifica (seja um indivíduo ou um grupo). A classificação que buscamos para os verbos do DISSE é uma classificação *consensual*; não é uma classificação *correta*, tampouco *definitiva*. Procuramos apenas alcançar um espaço em que haja concordância, em um nível que torne as categorias úteis para os profissionais da tradução. Procuramos um equilíbrio entre uma grande classe (todos os verbos agrupados em um grande grupo dos “verbos de elocução”) e os grupos de sentido muito específicos, propostos no primeiro teste. Quanto menos precisão (quer de uma classe, quer de uma definição), mais elasticidade; mais espaço para o consenso, mais possibilidades de inclusão. Decidimos, então, reduzir o número de grupos de sentido, mas tentando manter a informatividade necessária, para que não se perdesse o propósito e a utilidade da classificação.

Tomando essa perspectiva para a tarefa de organização do léxico, estávamos cientes de sua potencial inexequibilidade – que contrasta com a materialidade de sua utilização –, já alertadas por Austin (1962), em sua tentativa (frustrada) de classificação. Sobre os verbos que tornam explícita a força ilocucionária de uma declaração, Austin (1962, p. 149-150) afirma: “i distinguish five more general classes: but I am far from equally happy about all of them. [...] I call them classes of utterance, classified according to their illocutionary force”.

Como ilustração, apresentamos, no quadro 5, as classes propostas por Austin e os verbos incluídos em cada classe para, em seguida, analisar de forma breve a conclusão à qual Austin chega a respeito de sua classificação.

Quadro 5. Classes propostas por Austin (1962) para os verbos que tornam explícita a força ilocucionária

CLASSES PROPOSTAS E DEFINIÇÕES	EXEMPLOS DE VERBOS
(1) Vereditivos: exercício de julgamento.	reckon, assess, characterize, describe, analyze
(2) Exercitivos: afirmação de influência ou de exercício de poder.	appoint, order, command, recommend, veto, pray
(3) Comissivos: compromisso em assumir uma obrigação ou de declarar uma intenção.	promise, plan, intend, guarantee, vow, consent
(4) Comportamentais: adoção de uma atitude.	apologize, thank, welcome, congratulate, dare, criticize, approve
(5) Expositivos: esclarecimento de razões, argumentos e opiniões.	affirm, deny, report, remark, accept, inform, tell, ask, agree

Antes mesmo de propor a distribuição acima, Austin (1962, p. 124) se mostra insatisfeito com o resultado obtido. Mais adiante, tecendo alguns comentários a respeito das classes por ele estabelecidas, comenta:

As duas últimas classes [comportamentais e expositivos] são as que acho mais difíceis, e bem pode ser que não estejam nitidamente classificadas ou que estejam um tanto embaraçadas, ou mesmo que necessitem de uma classificação inteiramente distinta. Não estou, de modo algum, propondo nada definitivo. Os comportamentais criam problemas porque parecem demasiado heterogêneos, e os expositivos porque são excessivamente numerosos e importantes, e tanto parecem estar incluídos em outras classes quanto parecem, por vezes, ser *sui generis* de uma forma que não consegui esclarecer nem para mim mesmo. Bem se poderia dizer que todos os aspectos estão presentes em todas as classes. (tradução nossa)

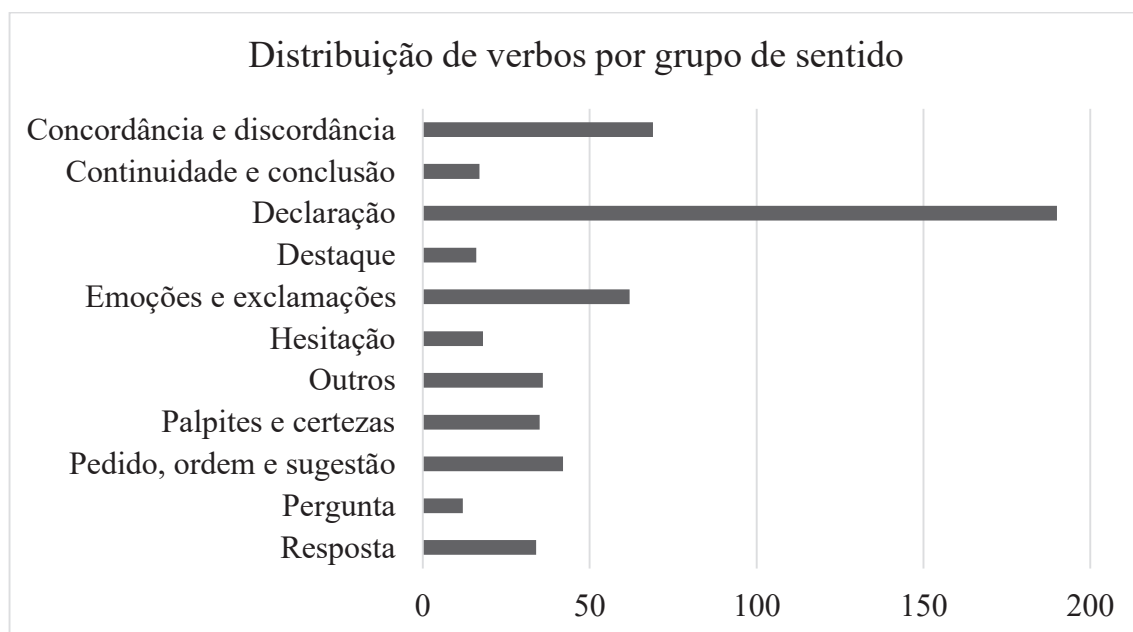
Assim, a tentativa, segundo Austin, fracassada de classificar esses verbos nos conforta quanto às complicações envolvidas na organização que propusemos. Entretanto, mantivemos nossa opção pela distribuição em grupos de sentido, mesmo cientes das dificuldades (ou da impossibilidade) de classificação de tais verbos, na medida em que esse tipo de classificação faz sentido *no contexto do léxico pretendido* (em oposição a uma tentativa de classificação geral desses verbos, por exemplo). Além disso, a ênfase em uma classificação motivada pelo sentido deve-se à sua utilidade para os usuários do DISSE – por isso não utilizamos uma classificação de viés sintático, menos sujeita à variação, como feita por Neves (2000). No momento da tradução, informações sintáticas a respeito de um verbo não costumam ser relevantes para o processo decisório do tradutor.

Segunda tentativa de classificação dos verbos de elocução

Passamos, então, à nova proposta de classificação dos verbos. Dessa vez, incluímos todos os verbos obtidos até então, que totalizam quase 300. Com base nessa nova lista, criamos onze classes, levando em conta as seguintes estratégias: (i) partimos dos resultados do primeiro teste, que acabaram por indicar alguns caminhos a respeito dos grupos; (ii) mesclamos alguns dos grupos que eram mais próximos em termos de sentido, para que o número total de grupos fosse reduzido; e (iii) levamos em conta a utilidade de cada grupo para o tradutor durante a sua prática. Analisamos as ocorrências dos verbos em contexto e os distribuímos nas novas classes propostas. A figura 2 apresenta as 11 classes e a distribuição dos verbos entre elas.

O grupo *declaração* foi proposto por dois respondentes para verbos que tinham sido enquadrados no grupo *afirmação*: os verbos *declarar* e *falar*. Consideramos que o nome *declaração* era mais adequado do que *afirmação*, pois o primeiro poderia abrigar um número maior de verbos (incluindo-se aqui, por exemplo, os verbos de negação). Os grupos *informação* e *opinião* também foram incorporados a esse novo grupo. Dessa forma, *declaração* constitui o grupo mais amplo, com 189 verbos dos 293 (64,7%). Muitos desses verbos também pertencem a outros grupos.

Os grupos *acordo* e *contestação* uniram-se para compor o grupo *concordância/discordância*. Entendemos que essa classe reúne verbos que indicam algum tipo de posicionamento em relação ao que o outro fala. A opção pela troca das palavras que denominam a classe se deveu à necessidade de

Figura 2. Distribuição de verbos por classe

Fonte: Freitas (2016).

maior clareza; com base nos resultados do primeiro teste, chegamos à conclusão de que os nomes *acordo* e *contestação* nem sempre ficaram claros para os participantes. Além disso, foi incorporada a essa classe a *censura*, pois consideramos *contestação* e *censura* campos semânticos afins.

Pedido e *sugestão*, grupos já existentes, juntaram-se em um novo grupo, que conta também com os verbos com ideia de *ordem*. A classe *acréscimo* recebeu um novo nome e teve seu escopo ampliado: *continuidade/conclusão*. Também ampliamos a abrangência da classe *exclamação*, nela incluindo os verbos que, de alguma forma, denotam emoção, sem necessariamente haver algum tipo de exclamação; temos, então, a classe *emoções/exclamações*. Por último, aumentamos o escopo da classe *hesitação*, que agora também abriga o que chamamos de *ponderação*.

A única classe realmente nova nessa segunda classificação é *palpites/certezas*. O único verbo que se enquadraria nesse grupo, *prometer*, estava antes classificado como *outros*. No entanto, considerando a nova lista (com muito mais verbos), essa classe foi ampliada, passando a incluir 35 verbos.

Apenas quatro grupos de sentido foram mantidos: *pergunta*, *resposta*, *destaque* e *outros*. Ainda que, comparativamente, os grupos *pergunta* e *resposta* englobem poucos verbos, decidimos por mantê-los pelo simples motivo de que o tradutor, quando busca uma tradução possível para *said*, costuma pensar, em primeiro lugar, se o contexto constitui uma declaração, uma pergunta ou uma resposta. Em geral, há pistas formais que facilitam esse pensamento, sendo a principal delas a pontuação.

Já o grupo *destaque* foi mantido, sem alteração de nome, pois, na primeira classificação, figurava entre os grupos com o maior número de verbos. Agora, trata-se da segundo menor grupo, com 16 ver-

bos, ficando atrás apenas de *pergunta*. Por fim, a manutenção do grupo *outros* se deve à necessidade de um grupo-coringa, que abrigue aqueles verbos que não se enquadram bem nos demais grupos, mas que, ao mesmo tempo, não justificam a criação de um grupo só para si.

Implementamos a segunda versão do teste no Rêve (Santos *et al.*, 2015), uma ferramenta *on-line* criada no âmbito da Gramateca (Santos, 2014) para auxiliar atividades de pesquisa linguística com base em *corpus* que tomam a anotação como processo de investigação. De fato, o processo de classificação de elementos linguísticos em contexto (que corresponde ao teste criado para validar as classes propostas) pode ser visto como um processo de anotação linguística, ainda que, no nosso caso, não tenha havido implementação. O uso da ferramenta trouxe facilidades tanto para a realização do teste por parte dos voluntários quanto para a análise dos resultados, uma vez que o Rêve produz um relatório de estatísticas das respostas ao teste.

Para esse teste, selecionamos 30 verbos apenas, pois julgamos que a versão anterior tinha ficado muito longa. Os critérios de seleção desses 30 verbos foram os seguintes: (1) quinze verbos correspondem a verbos do teste anterior, mais especificamente aos verbos que apresentaram mais discordância na classificação (pois receberam um número maior de grupos) e (2) os outros quinze foram retirados aleatoriamente dos onze grupos de sentido, porém nos certificamos de que havia pelo menos um verbo de cada grupo. O teste foi feito por 17 participantes.

A Figura 3 mostra a interface do teste aplicado e apresenta, também, a definição que demos a cada um dos grupos de sentido.

Figura 3. Interface da segunda versão do teste da classificação em grupos de sentido

Rêve Linguateca

Dizer Bianca

Atribua uma classe para o verbo em negrito, levando em conta o seu contexto na frase. É possível selecionar mais de uma classe, simultaneamente. Use a caixa para comentários e sugestões, se achar necessário.

DECLARAÇÃO: introduz qualquer tipo de declaração.
 PERGUNTA: introduz qualquer tipo de pergunta.
 RESPOSTA: introduz qualquer tipo de resposta a uma pergunta ou a uma declaração.
 DESTAQUE: qualifica o que foi dito como algum tipo de destaque ou realce.
 CONTINUIDADE/CONCLUSÃO: indica continuidade ou conclusão da fala do locutor.
 EMOÇÕES/EXCLAMAÇÕES: indica algum tipo de emoção ou exclamação por parte do locutor.
 ORDEM/PEDIDO/SUGESTÃO: indica um tom de ordem, pedido ou sugestão.
 HESITAÇÃO/PONDERAÇÃO: indica algum tipo de hesitação ou avaliação sobre o que foi dito.
 CONCORDÂNCIA/DISCORDÂNCIA: indica o posicionamento favorável ou contrário do locutor em relação a um interlocutor ou a um assunto em discussão.
 PALPITE/CERTEZA: indica palpite, incerteza ou certeza sobre o que foi dito.
 OUTROS: verbos que não se encaixam nas demais categorias.

Guardar

FSP940118-193: Os dois prestam medicina na Unesp, direito na Fuvest e engenharia de alimentos na Unicamp, prova que resolvem na mesma classe, a duas cadeiras de distância. «Fizemos inscrições juntos em todas as universidades», **conta** Grace.

FSP941118-105: Muita gente não gostou de vê-la no desfile», conta Walter Rodrigues. «Mas acho que, no fim, alguém tem que ser diferente», **conclui**. Bronie, como sempre, não estava nem aí.

DECLARAÇÃO, ☒ DECLARAÇÃO ☐ PALPITE/CERTEZA ☐ PERGUNTA ☐ RESPOSTA

Fonte: Freitas (2016, p. ?).

Apesar das reformulações, os resultados do teste não foram muito diferentes da primeira versão, corroborando as impressões de Austin. Dos 30 verbos testados, em apenas 19 (cerca de 63%) conseguimos uma concordância igual ou superior a 70% por parte dos respondentes (decidimos reduzir o número de 80% para 70% por conta do aumento do número de respondentes). O quadro 6 mostra os verbos cujas classes foram validadas.

Quadro 6. Lista dos verbos da segunda versão do teste, com as respectivas classes atribuídas

VERBO	CLASSES
contar	DECLARAÇÃO (94,1%)
concluir	CONTINUIDADE/CONCLUSÃO (82,4%)
queixar-se	EMOÇÕES/EXCLAMAÇÕES (70,6%)
prometer	DECLARAÇÃO (70,6%)
contrapor	CONCORDÂNCIA/DISCORDÂNCIA (82,4%)
prosseguir	CONTINUIDADE/CONCLUSÃO (100%)
espantar-se	EMOÇÕES/EXCLAMAÇÕES (100%)
anunciar	DECLARAÇÃO (100%)
esclarecer	DECLARAÇÃO (70,6%)
arrematar	CONTINUIDADE/CONCLUSÃO (76,5%)
continuar	CONTINUIDADE/CONCLUSÃO (100%)
convidar	ORDEM/PEDIDO/SUGESTÃO (70,6%)
censurar	CONCORDÂNCIA/DISCORDÂNCIA (82,4%)
frisar	DESTAQUE (94,1%)
emendar	CONTINUIDADE/CONCLUSÃO (100%)
festejar	EMOÇÕES/EXCLAMAÇÕES (94,1%)
apoiar	CONCORDÂNCIA/DISCORDÂNCIA (94,1%)
xingar	EMOÇÕES/EXCLAMAÇÕES (88,2%)
sondar	PERGUNTAR (82,4%)

Por outro lado, conseguimos reduzir a média de classes atribuídas por verbo, que ficou em 2,7. Os verbos *emendar* e *anunciar* receberam apenas uma classe, enquanto o verbo *advertir* teve o maior número de classes atribuídas: sete. O Quadro 7 mostra algumas das frases testadas e as classes atribuídas pelos participantes.

Desconsideramos classificações únicas, isto é, atribuídas por apenas uma ou duas pessoas. Assim como nos casos de anotação colaborativa, é possível que tais participantes, embora tenham feito o teste voluntariamente, tenham tido pouco “comprometimento com a tarefa”, fato já observado em Santos *et al.* (2015).

Uma das dificuldades do nosso teste, notada na primeira versão e confirmada na segunda, deve ser atribuída à natureza da tarefa. Percebemos que foi muito difícil para os participantes concentrar a classificação *apenas* no verbo. Em outras palavras, pensávamos, a princípio, que o verbo que intro-

Quadro 7. Linhas de concordância avaliadas na segunda versão do teste e que receberam o menor ou o maior número de classes

LINHAS DE CONCORDÂNCIA	CLASSES ATRIBUÍDAS
<p>Confessou a juíza ao condenar José à pena mais mínima que podia, 12 contos mais despesas de tribunal: «Às vezes a necessidade leva-nos a cometer coisas que são muito pouco aconselháveis», e depois emen- dou que «com certeza a sua filha prefere ter um pai que seja um bom exemplo em vez de ter um pai preso por crimes» e remendou que «inclusivamente pode estar a criar sentimentos de culpa à sua filha -- ‘o meu pai cometeu um crime por minha causa...’</p>	CONTINUIDADE/CONCLUSÃO
<p>O anúncio foi feito ontem pelo futuro presidente do Santos, Miguel Kodja Neto, e pelo prefeito David Capistrano (PT). Pelé esteve presente e anunciou que até sexta-feira deve divulgar a empresa que irá patrocinar o clube.</p>	DECLARAÇÃO
<p>«O resultado de tanta incapacidade, hesitação e má vontade será a ONU desacreditada, a Otan arruinada e os europeus desmoralizados. “ “ As relações entre Ocidente e o mundo muçulmano nunca mais serão as mesmas», advertiu o presidente. Disse ainda que a Rússia «está do lado dos agressores».</p>	DECLARAÇÃO EMOÇÕES/EXCLAMAÇÕES PALPITES/CERTEZAS HESITAÇÃO/PONDERAÇÃO ORDEM/PEDIDO/SUGESTÃO DESTAQUE OUTROS

duzia o relato condensava a carga semântica do que era dito. O teste nos mostrou, no entanto, que as pessoas se deixavam levar mais pelo que era dito do que pelo verbo empregado; a maioria das discordâncias do teste é fruto dessa influência do conteúdo relatado, como ilustra o trecho abaixo, classificado por um participante como “destaque”:

*O ala Dominique Wilkins, 34, deu uma entrevista coletiva em que se queixou dos dirigentes de sua ex-equipe, o Atlanta Hawks. «Eles não precisavam mostrar tanto desrespeito», **queixou-se** Wilkins, 34, trocado por Danny Manning, do Los Angeles Clippers*

Ainda assim, consideramos os resultados satisfatórios, e organizamos o léxico DISSE, mantendo os onze grupos de sentido que propusemos inicialmente. Além disso, dentro de cada grupo, listamos os verbos por ordem de frequência nos *corpora* consultados¹³. O tradutor, ao buscar um verbo para usar em sua tradução, terá, além do campo semântico, um critério relacionado à frequência de uso: ou estará em busca de um verbo mais comumente usado, ou então de um verbo mais heterodoxo. No primeiro

¹³ Na versão disponível online, também organizamos os verbos por ordem alfabética.

caso, o tradutor poderá direcionar o olhar para o início da lista, enquanto para o segundo caso, o fim da lista lhe dará opções mais interessantes, mais ricas de nuances. e o novo subtítulo.

Um olhar linguístico e linguístico-computacional para o DISSE

De uma perspectiva dos estudos linguísticos, é interessante notar a distribuição de verbos em cada grupo, o que pode também motivar o desenvolvimento de estudos contrastivos no que se refere aos verbos de elocução. De volta à Figura 2, vemos que, excluindo-se o grupo geral *declaração*, o grupo *emoções/exclamações* e o grupo *concordância/discordância* são aqueles mais densos, incluindo um grande número de verbos (78 e 72, respectivamente), o que pode sinalizar a frequência com que tais tipos de atitude aparecem associadas a verbos de elocução. De fato, quando buscamos pelos lemas dos verbos do grupo *emoções/exclamações* no serviço AC/DC, com o *corpus* Floresta, e indicamos na pesquisa que tais verbos precisam também pertencer ao campo semântico da emoção¹⁴, vemos que, dos 78 de nosso grupo inicial, 47 já estão incluídos neste campo semântico. Resta saber se os 31 verbos restantes – como por exemplo *gritar*; *bradar*; *desafiar*; *debater*; *provocar*; *exclamar* – também poderiam receber essa classe semântica, enriquecendo assim não apenas a anotação semântica do campo da emoção em português, mas, como consequência, estudos sobre a emoção na língua. O mesmo se aplica ao grupo *concordância/discordância*, que contém verbos como *reagir*, *alfinetar*, *reclamar*, *resignar-se* – e não à toa ambos os grupos compartilham verbos, como *ameaçar*, *desafiar*, *detonar*, *protestar*, *provocar*, *queixar(se)*. Aliás, deixando de lado o ponto de vista (das necessidades) do tradutor, vemos que os grupos *concordância/discordância* e o grupo *emoções/exclamações* poderiam se fundir em um grande grupo *emoções*, contribuindo para a anotação dessa classe mais ampla e que tem recebido bastante atenção no ambiente da linguística computacional (Pang e Lee, 2008; Santos e Maia, 2018).

Além do diálogo com a análise de emoções e sentimentos, o reconhecimento automático do discurso relatado interessa a, pelo menos, duas outras áreas que lidam com o processamento automático de textos: a tarefa de extração de citação (*Quotation Extraction*) (veja-se, por exemplo, Pareti *et al.* 2013) e a grande área das Humanidades Digitais (veja-se Santos *et al.*, 2018 para explorações em literatura de língua portuguesa). Em ambos os casos, para além da identificação automática do discurso relatado propriamente, a distribuição dos verbos de elocução em grupos semânticos, como a descrita aqui, viabiliza a investigação relativa à atitude de quem os profere, sendo responsável pela adição de mais uma camada de informação. Assim, na identificação das falas de personagens (sejam de obras literárias, um dos focos das Humanidades Digitais, ou de personagens de notícias, alvo tradicional da tarefa de identificação e extração de citação), a classificação mais fina com relação aos sentidos

¹⁴ A expressão de busca utilizada foi [sema="emo.*" & lema="continuar|manter|tornar|começar|concluir|acrescentar|retomar|completar|cortar|prosseguir|complementar|acrescer|rematar|reiterar|finalizar|emendar|arrematar|provocar|lembrar|assumir|discutir|entregar|cantar|atacar|temer|reagir|comemorar|contestar|interromper|gritar|disparar|animar|protestar|atirar|rir|ameaçar|confiar|divertir.*|chorar|desmentir|incentivar|homenagear|admirar.*|desafiar|queixar.*|desconfiar|rezar|sorrir|saudar|festejar|espantar.*|cumprimentar|entusiasmar.*|rebater|relembra|preconizar|desculpar|indignar.*|desabafar|censurar|orgulhar.*|detonar|clamar|empolgar.*|brindar|exclamar|berrar|xingar|alardear|confidenciar|gabar.*|alegrar.*|bradar|murmurar|resmungar|exultar|repreender|rememorar"]

dos verbos de elocução – e, por consequência, aos efeitos de sentido dos proferimentos/enunciados –, contribui para a caracterização de personagens em termos de agressividade, passividade e emoção: personagens que *gritam*, *protestam* e *bradam* (todos do grupo *emoções e exclamações*), por exemplo, em oposição a personagens que *consideram*, *avaliam*, *refletem* (grupo *hesitação e ponderação*).

Considerações Finais

Apresentamos aqui o processo de elaboração do DISSE, um léxico de verbos de elocução em português, construído para auxiliar profissionais de tradução, mas não apenas. O léxico contém verbos com usos atestados em contextos de elocução, nas variantes usadas no Brasil e em Portugal, e está público para consulta¹⁵.

A partir de um estudo anterior que elencou quase 300 verbos de elocução, nos propusemos a classificá-los em grupos de sentido, a fim de prover alguma organização e facilitar a consulta. Após duas tentativas e dois testes, distribuímos os verbos em 11 grupos. Sabemos que a classificação dos verbos de elocução em grupos de sentido é tarefa complexa – talvez impossível, como sugere Austin (1962). No entanto, tínhamos um problema concreto para resolver: organizar uma lista extensa de verbos, tendo em vista as necessidades dos tradutores.

Todo o material é público e está disponível à comunidade, o que pode incentivar a sua utilização em áreas que também reconhecem a relevância do dizer na língua, como a análise crítica do discurso, que recentemente vem se aproximando dos estudos com base em grandes *corpora*.

Referências

- AUSTIN, J.L. 1962. *Quando dizer é fazer*. Porto Alegre, Artes Médicas, 136 p.
- ARTSTEIN, R.; POESIO, M. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555-596. <https://doi.org/10.1162/coli.07-034-R2>
- BAKER, C.; FELLBAUM, C.; PASSONNEAU, R. 2017. Semantic Annotation of MASC. In: N. IDE; J. PUSTEJOVSKY (eds.). *Handbook of linguistic annotation*. Dordrecht, Springer Netherlands, p. 699-717. https://doi.org/10.1007/978-94-024-0881-2_25
- COSTA, B.F.S.; FREITAS, C. 2017. Verbos de elocução em português: um estudo descritivo com base em grandes corpora e motivado pela linguística computacional. *Fórum Linguístico*, Santa Catarina, 14(3):2266-2285. <https://doi.org/10.5007/1984-8412.2017v14n3p2266>
- COSTA, L.; SANTOS, D.; ROCHA, P. 2009. Estudando o português tal como é usado: o serviço AC/DC. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 7., 2009, São Paulo. *Anais...* São Paulo, STIL. Disponível em: <http://www.nilc.icmc.usp.br/til/>

¹⁵ Verbos organizados por ordem de frequência: <https://www.linguateca.pt/Gramateca/DISSEfrequencia.html>; verbos organizados por ordem alfabética: <https://www.linguateca.pt/Gramateca/DISSEalfabetica.html>

stil2009_English/Proceedings/stil/Costa-57572_1.pdf. Acesso em: 11/05/2019.

ELLIS, J. 1993. *Language, thought and logic*. Illinois, Northwestern University Press, 163 p.

FELLBAUM, C. 1998. *WordNet: an electronic lexical database*, Cambridge, MIT Press, 423 p. <https://doi.org/10.7551/mitpress/7287.001.0001>

FRANKENBERG-GARCIA, A.; SANTOS, D. 2002. Comparada, um corpus paralelo de português e de inglês na Web. *Cadernos de Tradução*, 9(1):61-79.

FREITAS, B. 2016. *O dizer em português: diálogos entre tradução, descrição e linguística computacional*. Rio de Janeiro. Dissertação (Mestrado). Programa de Pós-Graduação em Estudos da Linguagem, Pontifícia Universidade Católica do Rio de Janeiro, 133 p.

FREITAS, C. 2015. Corpus, linguística computacional e as humanidades digitais. In: M. LEITE; C. GABRIEL (eds.). *Linguagem, discurso, pesquisa e educação*. Petrópolis, De Petrus/FAPERJ, p. 18-46.

FREITAS, C.; ROCHA, P.; BICK, E. 2008. Um mundo novo na floresta sintá(c)tica - o treebank para português. *Calidoscópio*, São Leopoldo, 6(3):142-148. <https://doi.org/10.4013/cld.20083.03>

GARCIA, O.M. 2010. *Comunicação em prosa moderna: aprenda a escrever, aprendendo a pensar*. 27. ed. Rio de Janeiro, Editora da FGV, 548 p.

GARSDALE, R.; LEECH, G.; MCENERY, A. (eds.). 1997. *Corpus annotation: linguistic information from computer text corpora*. Michigan, Longman, 281 p.

KILGARRIFF, A. 1997. I don't believe in word senses. *Computers and the Humanities*, New York, 31(2):91-113. <https://doi.org/10.1023/A:1000583911091>

LANDES, L.; LEACOCK, L.; TENGI, R. 1998. Building semantic concordances. In: C. FELLBAUM (ed.), *WordNet: an electronic lexical database*, Cambridge, MIT Press, p. 199-216.

LEVIN, B. 1993. *English verb classes and alternations*. Chicago, The University of Chicago Press, 348 p.

MARTINS, H. 2000. Sobre a estabilidade do significado em Wittgenstein. *Veredas*, Juiz de Fora, 4(2):19-42.

NEVES, M.H.M. 2000. *Gramática de usos do português*. São Paulo, Editora da Unesp, 1037 p.

NIETZSCHE, F. 1873. Sobre a verdade e a mentira no sentido extra-moral. In: G. LEBRUN (ed.). *Obras incompletas*. São Paulo, Nova Cultural, p. 51-60.

OLIVEIRA, H.; PAIVA, V.; FREITAS, C.; RADEMAKER, A.; REAL, L.; SIMÕES, A. 2015. As Wordnets do Português. *OSLa - Oslo Studies in Language*, Oslo, 7(1):397-424.

PANG, B.; LEE, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, Boston, 2(1-2):1-135. <http://dx.doi.org/10.1561/15000000011>

PARETI, S.; O'KEEFE, T.; KONSTAS, I.; CURRAN, J. R.; KOPRINSKA, I. 2013. Automatically detecting and attributing indirect quotations. In: EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 10., 2013, Seattle. *Proceedings...* Seattle, EMNLP, p. 989–999. Disponível em: <https://www.aclweb.org/anthology/D13-1101>. Acesso em: 11/05/2019.

SANTOS, D. 2011. Linguatca's infrastructure for portuguese and how it allows the detailed study of language varieties. *OSLa - Oslo Studies in Language*, Oslon, **3**(2):113-128.

SANTOS, D. 2014. Gramateca: corpus-based grammar of Portuguese. In: J. BAPTISTA; N. MAMEDE; S. CANDEIAS; I. PARABONI; T. PARDO; M.G.V. NUNES (eds.). *International conference on computational processing of portuguese (PROPOR'2014)*, Cham, Springer International Publishing, p. 214-219.

SANTOS, D.; FREITAS, C.; LOPES, J.M. 2018. Comparando a literatura lusófona com outras literaturas: recursos para leitura a distância em português. In: S. HIGUCHI; C. RIBEIRO (eds.). *I Congresso Internacional em Humanidades Digitais no Rio de Janeiro (HdRio2018)*, Rio de Janeiro, CPDOC/FGV, 2018, p. 375-383.

SANTOS, D.; MAIA, B. 2018. Language, emotion, and the emotions: A computational introduction. *Language and Linguistics Compass*, Hoboken, **12**(5):1-17. <https://doi.org/10.1111/lnc3.12279>

SANTOS, D.; MARQUES, R.; FREITAS, C.; MOTA, C.; SIMÕES, A. 2015. Comparando anotações na gramateca: filosofia, ferramentas e exemplos. *Domínios da Linguagem*, Uberlândia, **9**(2):11-26. <https://doi.org/10.14393/DL18-v9n2a2015-2>

SANTOS, D.; ROCHA, P. 2005. The key to the first CLEF in Portuguese: topics, questions and answers in CHAVE. In: C. PETERS Carol et al. (eds.). *Multilingual information access for text, speech and images, 5th Workshop of the cross-language evaluation forum, CLEF 2004, Bath, UK, 2004*. Berlin/Heidelberg, Springer, Lecture Notes in Computer Science, p. 821-832.

WITTGENSTEIN, L. 1953. *Investigações filosóficas*. São Paulo, Abril Cultural, 226 p.

Submetido: 15/11/2018

Aceito: 13/05/2019