

# VERBOS DE ELOCUÇÃO EM PORTUGUÊS: UM ESTUDO DESCRITIVO COM BASE EM GRANDES *CORPORA* E MOTIVADO PELA LINGUÍSTICA COMPUTACIONAL

LOS VERBOS DECLARATIVOS EN PORTUGUÉS: UN ESTUDIO DESCRIPTIVO BASADO EN  
GRANDES CORPUS Y MOTIVADO POR LA LINGÜÍSTICA COMPUTACIONAL

REPORTED SPEECH VERBS IN PORTUGUESE: A CORPUS BASED DESCRIPTIVE STUDY  
MOTIVATED BY COMPUTATIONAL LINGUISTICS

Bianca Freitas Saburi Costa<sup>\*</sup>

Cláudia Freitas<sup>\*\*</sup>

Pontifícia Universidade Católica do Rio de Janeiro

RESUMO: Neste artigo, apresentamos os resultados e as etapas de um amplo estudo descritivo, com base em *corpora*, sobre os verbos de elocução em português. A motivação para o trabalho vem de uma tarefa da Linguística Computacional – a identificação de citação. Partimos, inicialmente, de traduções do verbo “said” em contextos literários para a apreensão de verbos de elocução. Em seguida, utilizando grandes *corpora* monolíngues, buscamos padrões léxico-gramaticais característicos dessa classe de verbos, a fim de ampliar a lista. Com a metodologia, foram identificados 293 verbos de elocução, distribuídos em oito padrões gerais típicos desse grupo de verbos. Os resultados são comparados com outros trabalhos em português, inglês e francês, que tematizam os verbos de elocução e atestam o sucesso da metodologia utilizada.

PALAVRAS-CHAVE: Verbos de elocução. Verbos *dicendi*. Descrição do português. *Corpora* monolíngues.

RESUMEN: En este artículo, presentamos los resultados y las etapas de un amplio estudio descriptivo, basado en corpus, sobre los verbos declarativos en portugués. La motivación para realizar el trabajo proviene de una de las tareas de la Lingüística Computacional: identificar las citas. Inicialmente, partimos de las traducciones del verbo “said” en contextos literarios para identificar los verbos declarativos. A continuación, mediante la utilización de grandes corpus monolingües, buscamos patrones léxico-gramaticales característicos de este tipo de verbos con el objetivo de ampliar la lista. Siguiendo esta metodología, se

---

<sup>\*</sup> Mestre em Estudos da Linguagem pela Pontifícia Universidade Católica do Rio de Janeiro. E-mail: bianca.saburi@gmail.com.

<sup>\*\*</sup> Doutora em Estudos da Linguagem pela Pontifícia Universidade Católica do Rio de Janeiro. Professora assistente do Departamento de Letras da PUC-Rio. E-mail: claudiafreitas@puc-rio.br.

identificaron 293 verbos declarativos, distribuidos en ocho patrones generales típicos de este tipo de verbos. Los resultados se comparan con los de otros trabajos realizados en portugués, inglés y francés, que abordan el tema de los verbos declarativos y demuestran el éxito de la metodología utilizada.

**PALABRAS CLAVE:** Verbos declarativos. *Verba dicendi*. Descripción del portugués. Corpus monolingüe.

**ABSTRACT:** In this article, we present the results as well as the procedures of a wide descriptive, corpus-based study on reporting verbs in Portuguese. The motivation for this research comes from a task carried out in Computational Linguistics – quotation extraction. In order to obtain reporting verbs used in Portuguese, our starting point was a list of translations of the verb “said” in literary contexts. Then, by using large monolingual corpora, we searched for lexical-grammatical patterns that characterize this verb class, in an attempt to broaden the list. Through this methodology, we identified 293 reporting verbs and distributed them amongst eight general patterns in which these verbs are typically found. The comparison between our research and other studies in Portuguese, English and French that investigate reporting verbs attests the success of the methodology that was developed.

**KEYWORDS:** Reporting verbs. *Verba dicendi*. Portuguese language description. Monolingual corpora.

## 1 APRESENTAÇÃO E MOTIVAÇÃO

Uma das características da linguagem humana é seu caráter dialógico e polifônico – boa parte de nossas atividades de linguagem envolve, em algum grau, a retomada da fala de outros. Essa retomada, que chamamos de discurso relatado, interessa também à Linguística Computacional/PLN (Processamento (automático) de Linguagem Natural), alinhando-se especificamente à tarefa de extração de citações (*quotation extraction*). O objetivo principal dessa tarefa é identificar as citações em um texto, relacionando-as a seus autores. O foco está na identificação de quem fala e no conteúdo dessa fala, e em geral privilegia-se o discurso direto, cujas marcas formais facilitam a detecção automática.

Uma vez que a estrutura do discurso relatado é relativamente regular na língua, abordagens que fazem uso de regras costumam ser bem-sucedidas. Por outro lado, as marcas formais que indicam a presença de uma citação, como aspas e travessão, não são exclusivas desse tipo de discurso, e por isso a relevância de verbos que indiquem a presença de um discurso relatado. A ocorrência abaixo, por exemplo, poderia ser falsamente detectada como discurso relatado, caso a única pista formal no desenvolvimento da tarefa fosse a pontuação:

- (a) O ataque do ombudsman ao vírus jornalístico apareceu na coluna «Chega de Ébola e Internet», de 28 de maio. Ademais, nem todo discurso relatado possui as marcas formais mencionadas, sendo estas específicas do discurso direto:
- (b) «Já fui mais nervoso», disse.
- (c) Maria de Fátima disse que planeja sair do Rio.

Justamente por ser de identificação mais difícil, a citação com discurso indireto nem sempre é abordada por sistemas de detecção automática, mesmo que estes constituam, segundo Paretí et al. (2013), quase metade dos discursos relatados.

Além disso, ainda que na tarefa de identificação de citação o foco não esteja exatamente no verbo, mas em quem fala e no conteúdo dessa fala, sabemos que a forma de introduzir o discurso relatado – isto é, o verbo escolhido – também é indicativa de posicionamentos. Desse modo, a existência de um léxico de verbos que indiquem a presença de discurso relatado, associados aos padrões léxico-sintáticos em que são usados, é um recurso de grande utilidade para o PLN em português.

O presente artigo tem como objetivo descrever o processo de criação de um léxico de verbos de elocução (doravante VE) com base em grandes corpora. Tais verbos são comumente conhecidos como “verbos *dicendi*” e correspondem também a “[...] verbos introdutórios de discurso (discurso direto ou discurso indireto)” (MOURA NEVES, 2000, p. 47).

Um estudo descritivo amplo dos verbos de elocução também pode interessar a outras áreas, como o ensino de português – quando se tematiza discurso direto e indireto – e a Pragmática, com destaque para o que é discutido em Austin (1962). Apesar de existirem

trabalhos sobre o tema, com uma perspectiva mais estilística ou pragmática, há uma carência de estudos descritivos, sobretudo com base em grandes corpora. A exceção é a *Gramática de usos do português*, de Maria Helena de Moura Neves (2000), que, com base em corpus, aborda os verbos de elocução, listando-os conforme sua presença em discursos direto e indireto. Nossos objetivos, neste trabalho, são diferentes – e os resultados, complementares. Ainda, os corpora que usamos, integrantes do projeto AC/DC (COSTA et al., 2009)<sup>1</sup>, estão disponíveis ao público, o que permite a outros confrontar ou continuar o trabalho apresentado aqui. Todos os exemplos apresentados neste artigo foram retirados dos corpora do projeto.

Muitos dos estudos que descrevem verbos, principalmente os voltados para a língua inglesa, tratam dos chamados “verbos de comunicação”, um grupo que tem abrangência muito maior do que os VE. Os verbos de comunicação correspondem a qualquer verbo relacionado à comunicação ou à fala, independentemente de haver algo sendo relatado. Como nosso principal interesse está no relato, essa nomenclatura e essa abrangência não nos pareceram proveitosas.

O restante do artigo se estrutura da seguinte maneira: na seção 2, apresentamos estudos que já trataram dessa classe de verbos, tanto no português quanto em outras línguas; na seção 3, descrevemos detalhadamente o processo de levantamento de verbos de elocução; e, na seção 4, tecemos algumas considerações finais.

## 2 DOS ESTUDOS DA LINGUAGEM À LINGUÍSTICA COMPUTACIONAL: UMA RESENHA DE LITERATURA SOBRE OS VERBOS DE ELOCUÇÃO

No Brasil, há um número considerável de estudos que tratam dos verbos de elocução. No entanto, a maioria desses trabalhos adota um viés mais estilístico, pragmático ou normativo, o que não significa que não tenham contribuído para o estudo dessa classe de verbos da língua portuguesa. Quando consideramos outras línguas que não o português, no caso específico dos verbos, destaca-se o trabalho de Beth Levin (1993), *English verb classes and alternations*.

Levin (1993) se baseia no princípio de que o comportamento sintático de um verbo é, em grande parte, determinado pelo seu significado, principalmente no que se refere à expressão e à interpretação dos argumentos dos verbos. Dessa forma, o comportamento sintático dos verbos poderia ser usado para investigar aspectos dos significados dessa classe de palavras (LEVIN, 1993). Levin (1993) apresenta 49 classes de verbos que são “sintaticamente relevantes” e “semanticamente coerentes”. Para cada classe, a autora fornece uma lista de verbos como exemplo, mas essas listas não têm a pretensão de exaustividade.

Entre as 49 classes de verbos propostas, destacamos a classe “verbs of communication”. Nessa classe, segundo Levin (1993), estão reunidos os verbos relacionados com a comunicação e a transferência de ideias. Tais verbos, por sua vez, são classificados em nove categorias semânticas, que têm relação com a fala, mas não necessariamente introduzem discurso.

A princípio, a subcategoria dos “verbos de dizer” (*say verbs*) corresponderia melhor aos nossos interesses, mas, ainda assim, os critérios que demarcam cada uma das subcategorias não são claros, e outras subcategorias parecem abrigar alguns dos verbos que são considerados, por nós, como VE. Entre os grupos que seriam considerados como VE estão os “verbos de reclamar”, compostos por verbos como *complain* e *object*, e os “verbos de aconselhar”, que incluem os verbos *advise* e *warn*). Em Freitas (2016), apresentamos e discutimos de forma mais detalhada as nove categorias de Levin (1993).

Levin (1993) deixa claro que seu tratamento dos verbos de comunicação é bastante breve, correspondendo basicamente aos verbos cujas propriedades não envolvem complementos sentenciais. Essa brevidade se deve ao fato de que o estudo sistemático dos complementos sentenciais de verbos não se encontra dentro do escopo do livro em questão. Apesar do escopo limitado em termos sintáticos, Levin (1993) consegue reunir 163 verbos de comunicação, e o trabalho por ela desenvolvido é uma importante referência não apenas na descrição, mas também no PLN, uma vez que a sua classificação embasa um dos grandes recursos léxicos usados no processamento automático das línguas, a VerbNet (KIPPER et al., 2006) e, por extensão, a sua versão brasileira, VerbNet-BR (SCARTON et al., 2012).

<sup>1</sup> As características de cada corpus explorado nesta pesquisa serão apresentadas ao longo do artigo, mais especificamente nas seções 3 e 4.

Diferentemente de Levin (1993), o presente estudo não exclui os complementos oracionais, que são extremamente produtivos na língua, principalmente no que se refere ao relato de discurso. Por outro lado, tratamos de um grupo de verbos bem mais restrito que os “verbos de comunicação” da autora.

Com relação à língua portuguesa, destacamos Othon M. Garcia (2010) e Maria Helena de Moura Neves (2000). O primeiro, um manual, adota um tom normativo em todo o seu conteúdo.

Garcia (2010, p. 149) define os verbos *dicendi* como aqueles “[...] cuja principal função é indicar o interlocutor que está com a palavra”. No que chama de uma “lista caótica”, Garcia (2010) relaciona cinquenta verbos. Além disso, o autor comenta o uso de verbos que, a princípio, não estão relacionados com a elocução:

Chegam mesmo, os mais imaginativos, a empregar verbos que nenhuma relação têm com a ideia de elocução, o que, do ponto de vista da sintaxe, poderia ser considerado como inadmissível, pois os *dicendi* deveriam ser, teoricamente pelo menos, transitivos ou admitir transitividade. (GARCIA, 2010, p. 149)

Em seguida, Garcia (2010, p. 150) introduz os verbos chamados *sentiendi*, “[...] que não são propriamente ‘de dizer’ mas ‘de sentir’, e que, por analogia, podem ser chamados *sentiendi* [...], [verbos] que expressam estado de espírito, reação psicológica de personagem, emoções”, entre outros. O autor explica que os verbos *sentiendi* são “vicários” dos *dicendi* e caracterizam as manifestações psíquicas dos personagens. Garcia (2010) afirma, entretanto, que alguns verbos que não admitem a transitividade devem, no discurso direto, vir antepostos à fala:

- (a) Mas João de Deus, vendo que Vasco não lhe dá atenção, *explode*:  
– Você pensa, seu Vasco, que estou disposto a aturar suas malcriações? (VERÍSSIMO apud GARCIA, 2010, p. 155)
- (b) – Você pensa, seu Vasco, que estou disposto a aturar suas malcriações? – *explodiu* João de Deus.

O exemplo (a), fornecido por Garcia (2010), seria a única possibilidade de uso do verbo *sentiendi* “explodir”, a princípio intransitivo, no discurso direto. Para ele, o exemplo (b) seria inadmissível. Para que um verbo como “explodir” seja aceitável na posição posposta, seria necessário que um verbo “legitimamente” *dicendi* o acompanhasse e, ainda assim, a ordem dos verbos teria de ser alterada, tal como no exemplo a seguir:

- (c) – O coitadinho tem andado aborrecido! – disse ela *lamentando-se*.

Apesar dessa restrição imposta aos verbos *sentiendi* por Garcia (2010), uma busca no corpus CHAVE (SANTOS; ROCHA, 2005), corpus monolíngue que integra o projeto AC/DC (COSTA et al., 2009), utilizado em nossa pesquisa, revela o emprego frequente do próprio verbo “lamentar” em posição posposta à fala descrita, como é possível observar nos exemplos a seguir:

- (d) «Pensei que conseguiria vencer e fiquei animado», *lamentou* Wilander.
- (e) «Qualquer dia eles me tiram», *lamenta*.
- (f) «Até hoje não obtivemos resposta», *lamenta* Pádua.

A pesquisa em corpus permite a observação de que nem sempre normas prescritas por gramáticas e manuais de redação se aplicam a dados reais. Em apenas um corpus, foram encontradas 392 ocorrências do verbo *lamentar* posposto a uma fala. Fica evidente, assim, que os dados reais relativizam, quiçá invalidam a limitação do uso dos verbos *sentiendi* estipulada por Garcia (2010), e com os dados é acentuado, também, o potencial do uso de corpus para este tipo de estudo.

O trabalho que mais se aproxima do nosso está em *A Gramática de usos do português*, de Maria Helena de Moura Neves (2000). A gramática foi desenvolvida a partir de uma base de dados de setenta milhões de ocorrências, armazenada no Centro de Estudos Lexicográficos da Unesp e composta por textos de literaturas “romanesca, técnica, oratória, jornalística e dramática”, o que, segundo

a autora, “[...] garante a diversidade de gêneros e permite a abrangência de diferentes situações de enunciação, incluindo a interação, sendo notável a representatividade da língua falada, encontrada na simulação que dela fazem as peças teatrais” (MOURA NEVES, 2000, p. 14).

No entanto, diferentemente do trabalho que relatamos aqui, o material consultado por Moura Neves (2000) não está disponível ao público, o que dificulta a reprodução das pesquisas feitas na gramática. Dentro do grupo de “verbos que têm complementos oracionais”, a autora apresenta os verbos de elocução.

Moura Neves (2000, p. 48) chama de “verbos de dizer” os “verbos de elocução propriamente ditos”, isto é, os “[...] verbos de ação cujo complemento direto é o conteúdo do que se diz”. Ao lado dos verbos de dizer, a autora propõe os “verbos que introduzem discurso, mas não necessariamente indicam atos de fala”, grupo este que também integra os VE.

Incluem-se no grupo dos verbos de dizer, também chamados pela autora de verbos *dicendi*, os verbos *falar* e *dizer*, que seriam “neutros”, juntamente com diversos outros verbos “[...] cujo significado traz, somado ao dizer básico, informações sobre o modo de realização do enunciado” (MOURA NEVES, 2000, p. 48), como os verbos *gritar*, *berrar* e *sussurrar*. Também seriam verbos de dizer aqueles que acrescem “noções sobre a cronologia discursiva”, como é o caso dos verbos *retrucar* e *repetir*.

Moura Neves (2000, p. 48) destaca que, entre os verbos de dizer, existem também aqueles verbos nos quais se encontra “[...] lexicalizado o modo que caracteriza esse dizer”. A autora cita como exemplos os verbos *queixar-se*, *comentar* e *responder*.

Uma vez que os VE se encontram inseridos no grupo dos verbos que têm complementos oracionais, Moura Neves (2000) oferece alguns quadros em que são analisados os tipos de complemento aceitos pelos VE e nos quais o discurso introduzido se apresenta. Contabilizamos 103 verbos compilados pela autora como VE.

No contexto do PLN, trazemos o trabalho de Sagot et al. (2010), que apresentam o processo de criação de um léxico de *quotation verbs*, tendo em vista o desenvolvimento de uma ferramenta automática de extração de citação.

A partir de uma lista prévia de 110 verbos, compilados em um estudo anterior, Sagot et al. (2010) analisaram um corpus de cinco mil notícias de jornais, procurando por estruturas que correspondessem à presença de citações diretas, indiretas e mistas. Após filtrar os resultados manualmente, chegaram a 836 configurações de citação associadas a verbos de citação. Com o desenvolvimento de estratégias para a depreensão dos verbos nessas estruturas, os autores obtiveram 232 verbos.

O caminho seguido por nós é, de certa forma, semelhante ao seguido por Sagot et al. (2010): com base em corpus, conduzimos um levantamento semiautomático dos VE do português. Analisamos as estruturas em que esses verbos mais comumente apareciam e, a partir daí, levantamos centenas de verbos empregados nessas configurações. A seção seguinte detalha este processo.

### 3 A TRADUÇÃO COMO PONTO DE PARTIDA: “SAID”

Para dar início à compilação dos VE, o primeiro passo foi realizar um levantamento dos verbos usados para introduzir a fala em português. Para tanto, escolhemos um material com farta presença de discurso relatado: textos literários. Assim, os dados iniciais foram obtidos por meio de consultas ao COMPARA (FRANKENBERG-GARCIA; SANTOS, 2002), um corpus paralelo bilíngüe inglês-português, composto por textos literários e suas respectivas traduções em inglês ou em português<sup>2</sup>, com anotação morfossintática, disponível on-line<sup>3</sup>.

<sup>2</sup> Vale ressaltar aqui que o corpus é composto por originais e traduções do português de Portugal, Angola, Moçambique e do Brasil. No entanto, considerando o escopo do nosso estudo, não julgamos relevante excluir nenhuma das variantes.

<sup>3</sup> Disponível em: <<http://www.linguateca.pt/COMPARA/index.php>>.

A opção pelo COMPARA foi duplamente motivada: por um lado, sabemos que nem sempre é óbvia a identificação de um dado verbo como introdutor de discurso relatado (por exemplo, *imaginar* ou *interromper*). Assim, a escolha pela exploração inicial no COMPARA da forma *said* buscou, também, garantir segurança quanto à seleção dos verbos: quando o texto original usa a forma *said*, não há dúvidas de que o verbo escolhido, em português, se refere a um discurso relatado:

TEXTO FONTE: «Don't,» I **said**, in a muffled voice.

TEXTO META: – Não! – **interrompi** numa voz abafada.

Adicionalmente, trata-se de um corpus composto por obras literárias editadas, alinhadas com suas respectivas traduções<sup>4</sup>. Com isso, temos mais confiança quanto à qualidade das traduções.

No COMPARA, buscamos traduções para a forma verbal *said* (no original). Usamos a expressão de busca [word="said"]<sup>5</sup>, com a restrição de resultados que correspondessem apenas a traduções do inglês para o português. A opção pela busca da forma *said* (em vez de *say*) se dá porque o verbo, na maioria das vezes, introduz discurso em sua forma pretérita, o que é confirmado por Biber et al. (1999).

A busca pelas traduções de *said* nos forneceu um material bastante rico para o nosso ponto de partida, sendo desnecessário recorrer a outras formas verbais de *say* ou a outros verbos equivalentes.

Como o COMPARA limita os resultados a mil linhas de concordância aleatórias, usamos exatamente essa amostra, apesar de o corpus apresentar 3.560 ocorrências de *said*. Foram analisadas, então, todas as linhas de concordância obtidas com a referida busca. A Figura 1 apresenta a interface do COMPARA e algumas linhas de concordância.

Concordância		
EBDLT1(87)	Being a patient I would normally have gone into the Abbey, the BUPA hospital near the cricket ground, but they had a bit of a bottleneck there at the time – they were refurbishing one of their operating theatres or something – and Nizar <b>said</b> he could fit me in quicker if I came into the General, where he works one day a week for the NHS.	Sendo um doente particular, era mais normal que tivesse ido para o Abbey, um hospital privado que ficava perto do campo de críquete; mas na altura havia lá uma confusão qualquer – estavam a arranjar uma das salas de operações ou coisa do género – e o médico disse que tinha possibilidades de me encaixar mais depressa se fosse para o hospital civil, onde trabalha uma vez por semana para os doentes da Segurança Social.
EBDLT1(99)	«Ward 3!» I <b>said</b> .	«Enfermaria 3 J.», respondi.
EBDLT1(123)	«That man says he has been here three days and nobody has taken any notice of him.» I <b>said</b> .	«Aquele homem diz que está cá há três dias e que ninguém ainda reparou nele », expliquei.
EBDLT1(124)	«Well, at least he's had some sleep.» <b>said</b> the house-doctor. «which is more than I've had for the last thirty-six hours.»	«Bem, pelo menos vai dormindo», disse a médica, «que é coisa que eu não faço há trinta e seis horas.»
EBDLT1(132)	I <b>said</b> , «I could walk, you know, in a dressing-gown.»	Disse-lhe que, mesmo em camisa de dormir, podia ir a pé.
EBDLT1(156)	«No, you've got to be wheeled.» he <b>said</b> .	«Não, tem de ir de maca », disse ele.
EBDLT1(167)	«Got a smoke on you?» <b>said</b> the nurse.	«Tens tabaco?», perguntou a enfermeira.
EBDLT1(175)	«I'll tell you a secret, though.» <b>said</b> Tom.	Mas vou contar-te um segredo », continuou o Tom.
EBDLT1(213)	«That's good, very good.» he <b>said</b> reassuringly.	«Isso é bom, muito bom », disse-me num tom encorajador.
EBDLT1(225)	«Jolly good.» he <b>said</b> , <i>spilling</i> ».	– Fantástico! – exclamou. – Colossal!

Figura 1: Interface da busca no COMPARA

Fonte: Frankenberg-Garcia e Santos (2002)

Ao longo da análise, eliminamos 42 ocorrências, pois não correspondiam à introdução de um relato, como exemplificamos abaixo:

ORIGINAL: «What they do, what Polly *said*, oh please let's Ange!»

TRADUÇÃO: – O que eles fazem, o que *disse* a Polly, por favor, Ange!

O Quadro 1 lista as demais 958 ocorrências de *said*, distribuídas pelo número decrescente de ocorrências. Os casos de omissão do verbo foram 53.

<sup>4</sup> Vale ressaltar que o COMPARA constitui uma reunião de obras literárias que são célebres a ponto de serem traduzidas para outras línguas. Todas as traduções foram publicadas por editoras (em oposição a traduções “livres” publicadas na internet, por exemplo). Isso significa que as traduções encontradas no corpus são consideradas válidas por profissionais da área, o que nos dá mais segurança para aceitar os verbos escolhidos para traduzir *said* como verbos de elocução. Entre os autores traduzidos para o português, estão Henry James, Lewis Carroll e Oscar Wilde.

<sup>5</sup> A expressão indica que queríamos apenas a palavra “said” nessa forma exata e mais nada, sem que se considerassem possíveis variações.

VERBO	Nº OCORR.	VERBO	Nº OCORR.	VERBO	Nº OCORR.
dizer	561	comunicar	3	anuir	1
responder	88	concluir	3	anunciar	1
perguntar	60	confessar	3	balbuciar	1
<i>omissão</i>	53	confirmar	3	berrar	1
comentar	20	falar	3	brindar	1
explicar	17	informar	3	censurar	1
acrescentar	9	observar	3	concordar	1
afirmar	9	retrucar	3	desabafar	1
contrapor	9	ripostar	3	escrever	1
exclamar	9	agradecer	2	espantar-se	1
retorquir	9	argumentar	2	indagar	1
continuar	8	contar	2	justificar-se	1
declarar	7	cumprimentar	2	lembrar	1
insistir	6	esclarecer	2	prometer	1
interromper	6	propor	2	saudar	1
prosseguir	6	queixar-se	2	sondar	1
repetir	5	querer saber	2	sublinhar	1
pedir	4	admirar-se	1	sugerir	1
replicar	4	admitir	1	tornar	1
assegurar	3	advertir	1	<b>TOTAL</b>	<b>958</b>

**Quadro 1:** Verbos usados como tradução de "said" no COMPARA, ordenados por frequência.

**Fonte:** Freitas (2016)

Apesar do número de ocorrências do verbo *dizer* ser significativamente maior do que o número dos demais verbos, correspondendo a 58,56% das traduções, é interessante notar que, em cerca de 40% dos casos, os tradutores optaram por não empregar o verbo *dizer*, fazendo uso de 58 verbos distintos, que incluem desde verbos frequentemente empregados como introdutores de discurso, como *perguntar* e *responder*, e outros verbos não tão comuns nessa posição, como *sondar*, *agradecer*, *brincar* etc., que raramente são listados como verbos de elocução, de modo que não comparecem em listas como as de Garcia (2010) e Moura Neves (2000).

A etapa seguinte teve como objetivo ampliar a lista de 58 verbos, por meio de uma estratégia semiautomática. Para tanto, foi feita a exploração, em momentos diferentes, de grandes corpora monolíngues do projeto AC/DC (COSTA et al., 2009): o CHAVE (SANTOS; ROCHA, 2005), de textos jornalísticos (98 milhões de palavras); o OBRAS, de obras de literatura brasileira disponíveis em domínio público (1,2 milhão de palavras), e o Floresta (FREITAS et al., 2008), majoritariamente jornalístico (seis milhões de palavras). Todo o material está público, na página do projeto AC/DC<sup>6</sup>.

#### 4 DEPRENSÃO DE PADRÕES DE USO

Dos 58 verbos iniciais, seis foram escolhidos para atuar como “sementes”, a fim de identificar *padrões* léxico-gramaticais tipicamente usados para introduzir um discurso relatado. Subjacente à identificação dos padrões de uso desses verbos, está a ideia de que os próprios padrões seriam capazes de facilitar a localização de demais VE no corpus. Os verbos escolhidos foram *dizer*, *perguntar*, *responder*, *admitir*, *contar* e *continuar*. Os três primeiros verbos foram selecionados por serem frequentemente referenciados como VE, e realmente foram os mais frequentes nas traduções (ver Quadro 1). Além disso, esses verbos são normalmente associados à classe de verbos *dicendi* em gramáticas e manuais de redação (como o de Othon M. Garcia) e, por isso, podemos considerá-los prototípicos.

Por outro lado, os verbos *admitir*, *contar* e *continuar* foram escolhidos por estarem associados a usos que nem sempre correspondem à elocução, como, por exemplo:

*O evento **contou** com a presença do arcebispo de Fortaleza, d. Aloísio Loscheider.*  
*Às 13h, a reportagem da Folha **contou** 83 pessoas na manifestação.*

Apesar de ser o maior corpus paralelo revisto de português e inglês do mundo, o COMPARA dispõe de apenas 723.807 palavras e está restrito a textos ficcionais, em geral romances. Por isso, para a observação dos verbos-semente em contexto e consequente depreensão de seus padrões léxico-sintáticos, recorremos a um corpus mais robusto: o CHAVE. Além disso, o fato de o CHAVE e o COMPARA serem de gêneros diferentes (jornalístico e literário, respectivamente) também contribuiu para essa escolha, uma vez que a variedade poderia fornecer uma visão mais abrangente do uso dessa classe de verbos.

A ampla frequência dos VE no CHAVE se, por um lado, corrobora o acerto na escolha do material, por outro, aponta para a inviabilidade da análise caso a caso de todas as ocorrências (mais de 350 mil), obrigando-nos a desenvolver estratégias para lidar com a grande quantidade de informações.

Deste modo, analisamos, primeiro, os três verbos que tipicamente indicam elocução: *dizer*, *perguntar* e *responder*. Lemos centenas de ocorrências desses verbos e avaliamos em que casos introduziam uma fala. A partir da análise do contexto desses verbos “gerais”, foi possível depreender oito padrões em que eram usados de maneira recorrente. Em seguida, verificamos se os demais verbos – *admitir*, *contar* e *continuar* – também eram usados com esses padrões.

A maneira mais comum de se descrever o discurso relatado é de acordo com o tipo de citação envolvida: citação direta, indireta e mista. A maior importância dada ao tipo de citação refere-se à relevância para a evidencialidade do que é relatado. No nosso caso, o interesse nos padrões (e não exatamente nos tipos de citação) é resultado de sua relevância para as buscas em corpora: os padrões – formalizações – correspondem a diferentes materializações das estruturas de elocução. Uma consequência da ênfase no tipo de citação é a ideia subjacente de que há uma relação direta entre um dado padrão e um dado tipo de citação. Por exemplo, o padrão VE + “que” invariavelmente se associaria ao discurso indireto. A observação dos contextos, no entanto, revelou que um mesmo padrão pode participar de diferentes tipos de citação. O Quadro 3 mostra a distribuição dos padrões por tipo de citação, e é interessante perceber como um mesmo padrão pode estar associado a diferentes formas de citar.

<sup>6</sup> A sigla AC/DC significa Acesso a Corpora/Disponibilização de Corpora.

PADRÃO	DIRETA	INDIRETA	MISTA	EXEMPLOS
1	x			<b>D:</b> «O aumento não vale para as diárias, só mensalistas », <b>afirmou</b> Paulo Octávio.
2	x			Até que uma amiga minha passou por ele e <b>disse</b> : «Oi, Fábio».
3	x	x	x	<b>D:</b> «Na França », <b>diz</b> ela, « a história é uma atividade muito prestigiada, e portanto muito masculina». <b>I:</b> Antes de tentar um acordo, ele passou em outra concessionária (não informou o nome), que, <b>disse</b> , teria lhe oferecido arcar com parte do II. <b>M:</b> A contribuição de Robin Williams a «Aladdin », <b>salienta</b> ele, « nenhum diretor, escritor ou animador poderia suprir».
4		x	x	<b>I:</b> Cauteloso, ele <b>disse</b> que não receberá empresários e empreiteiras. <b>M:</b> Ambas acabam por <b>reconhecer</b> que « roubos, sempre os houve», mas <b>acrescentam</b> que «agora é muito pior».
5	x	x	x	<b>D:</b> Quando os jornalistas perguntaram se se tratava da reeleição dele ou de Fidel, FHC respondeu, rindo: «Para ele, não é reeleição, ele quer a eternidade». <b>I:</b> O banco respondeu, informando-os que o depósito em causa tinha sido liquidado em 30 de Agosto de 1985, e que havia pago um cheque ao portador nesse montante, assinado pelo João. <b>M:</b> Vital Moreira <b>respondeu</b> , afirmando manter «tudo o que disse» e que «a comunicação social escreveu, sem que tenha havido desmentidos».
6		x	x	<b>I:</b> Ontem, Farouk Kaddoumi, dirigente do departamento político da OLP, se <b>disse</b> favorável a realização de uma nova conferência multilateral de paz, conforme proposta da Rússia. <b>M:</b> Ontem, na reapresentação na Vila Belmiro, ele se <b>disse</b> abatido, «principalmente por ter sido contra o São Paulo, de quem eu queria ganhar».
7	x	x	x	<b>D:</b> Quando os jornalistas perguntaram se se tratava da reeleição dele ou de Fidel, FHC respondeu, rindo: «Para ele, não é reeleição, ele quer a eternidade». <b>I:</b> Charles respondeu dizendo que não podia se envolver nessa questão. <b>M:</b> A acusação partiu de Miltiadis Evert, o líder da Nova Democracia, e o porta-voz do governo socialista, Telmahos Hyritis, <b>respondeu</b> acusando-o de «críticas irresponsáveis».
8	x	x	x	<b>D:</b> Ou ainda, como <b>escreveu</b> o também poeta Joseph Brodsky: «A verdadeira biografia de um poeta é quase idêntica à dos pássaros». <b>I:</b> O presidente Itamar Franco deve vetar o projeto, segundo <b>informou</b> o ministro da Fazenda, Fernando Henrique Cardoso. <b>M:</b> Se há «graves impropriedades» no relatório, como <b>afirma</b> o Planalto, elas devem decerto ser identificadas e contestadas.

**Quadro 2:** Distribuição dos padrões por tipo de citação.

Fonte: Freitas (2016)

A seguir, apresentamos detalhadamente os oito padrões, tendo em vista cada um dos seis verbos analisados. Para facilitar a leitura, optamos por já indicar ao longo da descrição dos padrões algumas análises.

### Descrição dos padrões

#### (1) Citação completa + VE:

(a) «*Ninguém da minha família participa de sequestro*», **disse** Silva.

Com citações diretas, é fácil identificar os argumentos da citação nesse padrão. A pontuação característica desse padrão seria um dos facilitadores na identificação desses argumentos. No entanto, vale ressaltar que a pontuação nem sempre está presente:

(b) *Os carros estavam saindo de traseira*, **explicou** Keith Wiggins, o chefe da equipe.

O exemplo (b) constitui um caso em que o discurso é relatado de forma direta e sem o auxílio de pontuação típica de diálogos (aspas ou travessão). Ainda sobre a pontuação, as aspas costumam figurar em textos jornalísticos, enquanto em outros gêneros textuais, como, por exemplo, no texto literário, as aspas são frequentemente substituídas por travessões, de acordo com as orientações dadas por editoras. No caso das citações com travessão, a vírgula costuma ser omitida, segundo as regras ditadas pelos manuais de redação. Garcia (2010), por exemplo, afirma que “[...] o travessão torna prescindível qualquer outro sinal de pontuação, salvo os pontos de interrogação, de exclamação e as reticências”. O exemplo abaixo, que pertence à amostra de mil ocorrências retiradas do COMPARA, ilustra esse uso do travessão:

(a) – *E como podemos fazer uma nova série sem Debbie?* – eu **disse**.

Este padrão foi o primeiro a ser identificado na análise das ocorrências do verbo *dizer*, mas também aparece com todos os outros verbos, o que indica sua produtividade. A partir dele, portanto, podemos ter mais chances de encontrar outros VE. Essa hipótese se confirmará mais adiante.

#### (2) VE + citação completa:

(a) Zico, em conversa exclusiva com a Folha, **disse**: «*Estou muito contente com essa homenagem*».

(b) Uma hora após Gaviria ter evitado responder de forma categórica à pergunta, a Folha **perguntou** a Fujimori: «*Houve transparência nas eleições peruanas?*»

Pertencente ao discurso direto, tal como o Padrão (1), este também é um padrão extremamente corrente no português. A pontuação comumente empregada neste padrão (dois-pontos, travessões e aspas), tal como no Padrão (1), pode contribuir para a facilidade na identificação dos argumentos. Aqui, as aspas também podem ser substituídas por travessões, como mostra o exemplo (c):

(c) *Ao ouvir isto, Sofia virou-se para sua mãe, e empunhando a escova de lavar pratos*, **disse**:  
– *Saiba que ele é perfeitamente bom da cabeça*.

Tal como no anterior, o Padrão (2) foi inicialmente encontrado nas ocorrências de *dizer*, para, em seguida, ser fartamente encontrado com os demais verbos.

#### (3) Citação intercalada por VE

(a) «*Na Califórnia – diz ela – é cada vez maior o número de pessoas que come pouca carne ou que se tornam vegetarianas*».

(b) Como, **perguntará** o leitor, o execrando Tratado de Maastricht prevê uma coisa dessas?

- (c) «Vossa Santidade», **respondi**, «não só eu estou contente, todos nós estamos muito contentes».
- (d) «Sem dúvida alguma, ele é o melhor boxeador do mundo!», **admite**, «mas enquanto ele continuar dizendo que é o melhor lutador, estou pronto a desafiá-lo».
- (e) O helicóptero do líder do ANC foi apedrejado por manifestantes do Inkhata e a própria residência do rei foi atacada, **contaram** os seus conselheiros, obrigando o soberano a procurar refúgio numa quinta nos arredores.
- (f) Em 93, **continua**, a produção mundial foi de 18,7 milhões de toneladas e o consumo mundial, de 19,4 milhões.

O Padrão (3) admite os três tipos de citação e pode ser encontrado em textos tanto jornalísticos quanto literários. Nesse padrão, não é obrigatória a explicitação de todos os argumentos, mais especificamente do agente. Além disso, é bastante comum que o sujeito da ação esteja em um período anterior.

#### (4) VE + citação introduzida por conjunção subordinativa

- (a) Sobre a lateral esquerda, Parreira **disse** que quem deve jogar amanhã é Leonardo.
- (b) Me **perguntaram** se esse era o momento oportuno.
- (c) Genro, em tom duro também, **respondeu** que aquela não era uma visão de um dirigente de expressão nacional como Dirceu.
- (d) O teólogo católico **admite**, no entanto, que o tema das mulheres é um problema real, ao qual a Igreja Católica também tem que dar resposta.
- (e) FHC **contou** aos parlamentares que coube ao próprio Arida a indicação de Loyola.
- (f) Mas ante um persistente e interrogativo olhar do administrador, como que a intimá-lo a falar, ele **continuou** que não lhe tinha emprestado as pedras por uma questão de confiança, porque neste aspecto até poderia dizer que não confiava nem nos seus próprios dentes porque de quando em vez lhe mordiam a língua...

O Padrão (4) é rígido em termos de deslocamento sintático. O autor da citação ocupa o lugar convencional de sujeito, antecedendo a oração subordinada. Na linha de concordância (a), por exemplo, temos a citação, na sua forma indireta (“quem deve jogar amanhã é Leonardo”), a ação (“disse”) e o autor da citação (“Parreira”). Devemos ressaltar, entretanto, que a presença do autor da citação não é obrigatória; nesse caso, é possível encontrar o referente do sujeito em um período anterior.

Trata-se de uma variação do Padrão (4), com a adequação ao discurso direto (inclusão de dois-pontos):

- (g) **Acrescentou** ainda que: «A prazo, o financiamento do Ensino Superior deverá crescer, a par do aumento da sua frequência e dos resultados alcançados».

O padrão (4) é bastante recorrente na língua portuguesa, por conta da ampla gama de verbos que podem ser empregados dessa forma. No entanto, é justamente essa variedade que dificulta o nosso trabalho de identificar VE, pois essa estrutura não é necessariamente associada ao relato de discurso:

- (h) Pessoalmente, **considero** que Mariano Gago foi, de longe, o melhor ministro da era «democrática».

Na linha de concordância (h), teríamos, a princípio, um indício de citação mista. Contudo, o verbo *considerar*, tal como os verbos *admitir*, *contar* e *continuar*, tem outros sentidos que não correspondem ao de VE. No exemplo (h), o fato do verbo estar no presente e na primeira pessoa do singular já aparece como um impedimento para o relato, pois não julgamos ser possível relatar algo que ainda não foi dito. Justamente por conta do tempo verbal e da primeira pessoa, acreditamos que, em (h), o verbo expressa o sentido de “definir-se, depois de reflexão, sobre (determinada coisa); julgar” (HOUAISS, 2016). Além disso, temos que uma das marcas formais típicas de diálogos, as aspas, não cumprem apenas essa função. Muitas vezes, usamos as aspas apenas como recurso de ênfase ou destaque, como acreditamos ser o caso acima.

A oração subordinada substantiva objetiva direta que compõe o padrão (4) também é possível em sua forma incompleta, com as respostas “sim” e “não”:

(i) *Nem o governo, nos momentos de maior sinceridade, seria capaz de **responder** que sim.*

O verbo *perguntar* apresenta-se neste padrão de forma diferente dos demais verbos, assemelhando-se, por sua vez, a outros verbos que denotam pergunta. Como corresponde a uma pergunta convertida para o discurso indireto, encontramos comumente a conjunção integrante “se” no lugar de “que”, tal como na linha de concordância (b), além de outras conjunções integrantes e de pronomes interrogativos:

(j) ***Perguntei** por que eles queriam que justo a Executive Outcomes fizesse isso.*

(k) *Tivemos o desabafo do presidente da República quando um jornalista **perguntou** o que ele faria se dependesse do salário mínimo.*

#### (5) VE + citação em oração reduzida de infinitivo

(a) *Um dia antes de disputar a eleição, Sanguinetti **disse** achar que seu país está chegando ao Mercosul em desvantagem com os demais.*

(b) *Sem muita convicção, um deles **respondeu** saber de tal urgência.*

(c) *O ministro russo de Cultura, Evguêni Sidorov, **admitiu** ter tocado no tesouro.*

(d) ***Contou** ter feito a campanha de lançamento na Argentina da marca Cica.*

Em uma busca no corpus CHAVE, obtivemos 870 ocorrências de VE com complemento no infinitivo. Encontramos 73 verbos diferentes na posição de complemento.

No que se refere aos argumentos, o Padrão (5) apresenta a mesma estrutura e o mesmo processo de identificação da citação, do autor e do verbo introdutor de discurso relatado mostrados no padrão (4). A rigidez no deslocamento dos argumentos também se faz presente neste padrão, o que facilita o processo de identificação.

Foram encontradas apenas 21 ocorrências do verbo *continuar* seguido de infinitivo<sup>7</sup>. No entanto, em todas elas, *continuar* tem o sentido de “persistir”. É bem incomum que o complemento do verbo *continuar*, quando se trata de um infinitivo, venha desacompanhado da preposição “a”, como na linha de concordância (e):

(e) *Tirar um curso superior em Portugal **continua** depender das condições económicas.*

(f) *O grupo dos times pequenos **continua** a existir.*

#### (6) VE + pronome pessoal oblíquo + adjetivo ou particípio

(a) *O governo do México se **disse** disposto a negociar uma trégua.*

Esse é um padrão relativamente frequente do verbo *dizer*. Contudo, é difícil determinar se o verbo, nesse padrão, está sendo usado como VE, uma vez que tem como intuito atribuir característica ao autor da suposta citação. Na frase (a), o governo do México (autor) considerou que estava disposto a negociar uma trégua, ou seja, o governo do México atribuiu a si a característica “disposto”.

(b) *Ele se **disse** traído por Viljoen.*

<sup>7</sup> Para tanto, foi usada a expressão [pos="V.\*" & lema="continuar"]@[pos="V" & temcagr="INF"].

No exemplo (b), que difere de (a) por seu complemento ser um particípio e não um adjetivo<sup>8</sup>, observamos a mesma limitação quanto ao intuito do locutor. As questões são: “traído por Viljoen” é a coisa dita? Tem-se aqui, de fato, uma citação? O fato de que pouquíssimos verbos são empregados nessa estrutura – apenas cinco, como apresentamos no quadro 4 – já minimiza as chances do Padrão (6) ser um bom indicio de que temos um VE. Entre os seis verbos analisados, apenas *dizer* apresentou ocorrências que se enquadram no padrão e, em uma busca posterior, obtivemos apenas cinco verbos empregados no Padrão (6):

LEMA	OCORRÊNCIAS
dizer	26
declarar	6
revelar	3
reconhecer	2
defender	1

Quadro 3: Verbos encontrados dentro do Padrão (6).

Fonte: Freitas (2016)

### (7) Citação + VE + oração reduzida de gerúndio

O verbo *dizer* não foi encontrado introduzindo discurso com esse padrão. As linhas de concordância obtidas corroboram a teoria de que a oração reduzida de gerúndio acrescenta uma informação nova ao período, em vez de simplesmente reiterar o verbo de dizer que a antecede:

(a) «É, na época eu era aluno de pós-graduação», **diz** brincando.

Assim como o verbo *dizer*, o verbo *perguntar* não foi encontrado nesse padrão. As ocorrências de *perguntar* nessa estrutura correspondem a outros usos das orações reduzidas de gerúndio, como, por exemplo, para indicar ações simultâneas ou consecutivas:

(b) “Este aqui?”, **perguntou**, apontando para um homem moreno, cheio de brilhantina no cabelo e bigodinho à Errol Flynn.

(c) “Quem recebe é corrupto”, **perguntou**, provocando risos na plateia.

Os verbos *responder* e *admitir* são outros casos que deixam dúvidas. Por exemplo, seria a oração reduzida uma caracterização do *responder*? Seria uma ação adicional à ação de responder? As ocorrências encontradas para esses dois verbos não parecem validar o Padrão (7) como um forte candidato a padrão de uso dos VE.

(d) A hierarquia católica costuma **responder** dizendo que a igreja não é uma democracia.

(e) Mais modesto é, sem dúvida, o balanço em relação à evolução conseguida nas taxas de juro: «tem vindo a verificar-se uma evolução no sentido favorável, embora permaneçam no mercado de crédito taxas de juro reais elevadas», **admite**, sublinhando no texto entregue aos jornalistas que «a descida das taxas de juro não tem sido uniforme nos vários segmentos (dívida pública, crédito a grandes empresas, «PME Prestígio» e outras PME e particulares)».

<sup>8</sup> A classificação das formas participiais é bastante controversa na literatura linguística, como demonstra Trugo (2016).

Nas poucas ocorrências de oração reduzida de gerúndio após o verbo *contar*, o verbo da oração reduzida não auxilia na introdução de discurso, mas sim atribui algum tipo de característica ou especificidade ao ato de contar, tal como se vê nos exemplos abaixo:

(f) *Ele relata as histórias que lhe **contaram**, usando palavras que mostram que ele sabia aquilo a que se referia ao escrever.*

Não foram encontradas ocorrências do verbo *continuar* sendo empregado como introdutor de discurso relatado e seguindo o Padrão (7). Em todas as ocorrências em que é sucedido por um verbo no gerúndio, *continuar* apresenta o sentido de “prosseguir”:

(g) *Disse que reconhecia nele a virtude de **continuar** dizendo o que fez.*

#### (8) VE em oração subordinada adverbial conformativa + citação

(a) *E como **diz** o locutor Fiori Giglioti em suas transmissões de futebol, «o tempo passa, torcida brasileira».*

(b) *Então, como **perguntou** um professor presente no Curso de Verão, «não será necessário mudar o essencial da relação pedagógica, da filosofia do ensino e da organização da escola?»*

(c) *«Boa pergunta!», como **responderia** ele.*

(d) *Coca-Cola, eles provaram pela primeira vez no navio Merida, como **contou** à Folha o frentista Hugo Berto Isquierdo, 22.*

(e) *«O reforço das acções necessários» para a prossecução desses objectivos de luta contra a evasão e a fraude é, como **continuou**, «uma das prioridades da DGCI» e para o qual tanto este departamento como a Inspeção-Geral de Finanças eram «organismos vitais».*

Esse padrão caracteriza um caso bem peculiar dos VE no que se refere aos argumentos. Nos exemplos acima, pode-se ver que o que temos não é o simples relato de um discurso, mas a apropriação de uma fala proferida anteriormente. Nesse padrão, temos acesso ao autor “original” da citação, mas não necessariamente sabemos quem é o locutor no momento da enunciação, pois é ele próprio quem faz o relato.

Em contraste com os sete padrões anteriores, o Padrão (8) dispõe de uma estrutura sintática um pouco mais rígida. Mais especificamente, nesse padrão, não é possível omitir o sujeito do verbo introdutor de discurso relatado. A explicação provavelmente está fundada na peculiaridade da apropriação de uma fala anterior pelo locutor da frase; como o “dono original” da fala não é o sujeito do período que a introduz no relato, torna-se obrigatório explicitá-lo.

Abaixo encontram-se algumas variações desse padrão, que é observado em todos os tipos de discurso:

(a) *A América, como **dizia** Miles Davis, é um país maravilhoso onde você pode reunir mil vozes para gravar um unísono.*

(b) *A resposta, meu amigo, é sussurrada pelos ventos, como **diria** a canção.*

(c) *E como **disse** o pagodeiro na MTV: tudo na vida é passageiro, menos o cobrador e o motorista.*

(d) *Até chegar a uma «produtora estruturada», como **diz** Sandra, foram «meses de angústia».*

Os verbos *perguntar*, *responder* e *continuar*, segundo as nossas pesquisas, não costumam ser empregados no Padrão (8), apresentando três, duas e uma ocorrências, respectivamente. Apesar de os números não serem muito altos, a possibilidade de emprego dos seis verbos no padrão o torna um bom candidato para nos ajudar a levantar os verbos introdutores de discurso relatado.

### Análise da produtividade dos padrões encontrados

A análise da seção anterior está esquematizada na Tabela 1, que mostra quais padrões foram aceitos por cada verbo analisado:

PADRÃO VERBO	1	2	3	4	5	6	7	8
DIZER	sim	sim	sim	sim	sim	sim	<b>não</b>	sim
PERGUNTAR	sim	sim	sim	sim	<b>não</b>	<b>não</b>	<b>não</b>	sim
RESPONDER	sim	sim	sim	sim	sim	<b>não</b>	sim	sim
ADMITIR	sim	sim	sim	sim	sim	<b>não</b>	sim	sim
CONTAR	sim	sim	sim	sim	sim	<b>não</b>	<b>não</b>	sim
CONTINUAR	sim	sim	sim	sim	<b>não</b>	<b>não</b>	<b>não</b>	sim

**Tabela 1:** Ocorrências dos seis verbos em cada padrão analisado.

Fonte: Freitas (2016)

Com base na Tabela 1, é possível esboçar duas conclusões. A primeira delas é que todos os verbos analisados, por serem aplicáveis na maioria dos padrões estabelecidos, podem ser classificados como VE. Ainda que o verbo *continuar* não tenha sido encontrado em três dos oito padrões, isso não o exclui da classe de verbos, principalmente porque, na verdade, talvez devam ser considerados apenas cinco padrões – o que nos leva à segunda conclusão.

Como foi dito anteriormente, os padrões gramaticais utilizados com os VE seriam propostos a partir da análise dos seis verbos “semente” e, simultaneamente, esses mesmos seis verbos colocariam os padrões encontrados à prova. Uma vez que somente o verbo *dizer* é aplicável ao Padrão (6) e só os verbos *responder* e *admitir* se encaixaram satisfatoriamente no Padrão (7), a generalidade desses padrões é questionável e estes não nos ajudam muito a identificar outros VE. Por isso, daqui em diante, esses padrões serão deixados de lado.

O Padrão (5) foi aceito por quatro dos seis verbos, o que, para nós, é uma sugestão de que o padrão pode servir para introduzir discurso relatado. No entanto, a sua pouca precisão levaria, inevitavelmente, à recuperação de inúmeros outros verbos como candidatos a VE, mesmo sem sê-lo. Por isso, decidimos desconsiderá-lo. De fato, tínhamos outros cinco que serviam melhor ao nosso propósito: construir um léxico dos VE. Considerando os cinco principais padrões, passamos à etapa seguinte, de ampliação da lista de verbos.

## Ampliação da lista de VE: a construção de um léxico

Com base nos padrões, criamos as expressões para buscas na interface do AC/DC, a fim de ampliar o léxico dos verbos. As expressões estão listadas em Freitas (2016).

Logo percebemos que os padrões, sozinhos, não garantiriam a presença de um VE, uma vez que recuperavam estruturas como os exemplos abaixo, indicando a necessidade de outra análise caso a caso:

- (a) *Verificada a licitude da «pré-campanha », **regressemos** ao caso do leitor.*
- (b) *Ao longo de quase dois anos, os habitantes de Reveles ainda **acreditaram** que os trabalhos de estabilização dos solos pudessem vir a evitar o pior.*

Foi o que fizemos, desta vez com um corpus de tamanho inferior ao CHAVE, mas ainda assim de grandes dimensões: usamos o corpus Floresta (FREITAS et al., 2008), composto por 6.046.541 palavras.

Para cada expressão de busca usada, pedimos que nos fosse fornecida a distribuição dos verbos por lema, como ilustra a Figura 2:

### Resultados da procura

Fri Aug 18 15:43:25 WEST 2017

Procura: [word="","",\*|>|<|<|>"] [word=","] @[pos="V,\*" & temcagr!=",\*PCP,\*GER,\*PASSIVA,\*FUT\_SUBJ,\*PR\_SUBJ,\*FUT\_IND,\*0|,\*PR\_PROG\_IND,\*IMPF\_FUTHAVER\_IND,\*IMPF\_SUBJ,\*COND\_IND,\*"]  
Distribuição de lema

Corpo: Floresta Sintá(c)tica v. 2.5

8040 casos.

#### Distribuição

Houve **581** valores diferentes de lema.

dizer	2318
afirmar	913
explicar	468
contar	365
ser	222
comentar	159
lembrar	124
revelar	108
acrescentar	91
declarar	90
concluir	83
completar	81
perguntar	77
garantir	71
brincar	64
responder	61
referir	55
observar	54
escrever	53
falar	52
lamentar	51

Figura 2: Distribuição dos lemas conforme o padrão, consultando o corpus Floresta.

Fonte: produzido pelas autoras

Para que a análise fosse viável, não consideramos verbos com apenas uma ocorrência, e dispensamos da leitura caso a caso os 58 verbos já validados pela primeira etapa.

O resultado desta etapa foi uma lista de 293 verbos. O Quadro 5 apresenta a distribuição dos verbos segundo os padrões de busca.

Os dados incluem repetições, uma vez que um mesmo verbo pode participar de diferentes padrões. A lista completa dos verbos encontrados está disponível na página eletrônica da Gramateca (SANTOS, 2014), na seção “O relato em português”, por ordem de frequência e por ordem alfabética.

PADRÃO	VERBOS INTRODUTORES DE DISCURSO RELATADO
Padrão 1	219
Padrão 2	107
Padrão 3	137
Padrão 4	156
Padrão 8	52
TOTAL	293 (sem repetições)

**Quadro 4:** Número de verbos encontrados com as expressões de busca em cada padrão.

**Fonte:** Freitas (2016)

Com base no Quadro 4, é possível tirar algumas conclusões. Em primeiro lugar, vemos que o Padrão (1) é o mais produtivo de todos, uma vez que a porcentagem de verbos encontrados nesse padrão em relação ao total de 293 é a maior de todas: 74,74%. O Padrão (8), por sua vez, foi o menos produtivo, apresentando apenas 17,74% dos verbos compilados.

### Comparação dos resultados

Apresentamos aqui os resultados que obtivemos com a identificação dos padrões, comparando-os, sempre que possível, com os trabalhos já mencionados sobre o *dizer*. O Quadro 6 mostra a quantidade de VE encontrados ou listados por trabalhos citados nesta pesquisa:

PUBLICAÇÃO	IDIOMA	CLASSE DE VERBOS ANALISADA	QUANTIDADE DE VERBOS
Sagot et al. (2010)	francês	verbos de elocução	232
Levin (1993)	inglês	verbos de comunicação	163
Garcia (2010)	português	verbos de elocução	77
Moura Neves (2000)	português	verbos de elocução	103
<b>Este trabalho</b>	<b>português</b>	<b>verbos de elocução</b>	<b>293</b>

**Quadro 5:** Quadro comparativo dos verbos de elocução levantados em diversos estudos.

**Fonte:** Freitas (2016)

O Quadro 5 evidencia o sucesso da abordagem que adotamos nesta pesquisa. No entanto, os números, por si só, podem dar a entender que todos os autores tratam dos mesmos fenômenos, com os mesmos interesses, o que não é o caso. A terceira coluna do quadro mostra qual foi a classe de verbos abordada por cada um dos trabalhos, de acordo com a nossa classificação.

Sagot et al. (2010) têm interesse na extração automática de citação, e seu interesse nos verbos refere-se exclusivamente a resolver a tarefa pretendida; Biber et al. (1999) e Levin (1993) são trabalhos descritivos, que têm o inglês como língua de interesse e consideram a classe mais ampla dos *verbos de comunicação*; Garcia (2010) e o *Manual de Redação e Estilo do Estado* (MARTINS, 2016) são manuais de redação; e Moura Neves (2000) refere-se a uma gramática de usos do português.

Moura Neves (2000), tal como já afirmamos anteriormente, compilou 103 VE, em contraste com os 293 verbos desta pesquisa. Entre os 103 verbos da *Gramática de usos do português* (MOURA NEVES, 2000), não estavam presentes 25 na lista de verbos que reunimos. Os 25 verbos são: **acalmar**, **agastar-se**, **aguilhoar**, antecipar, boquejar, **bronquear**, **bulir**, **caçoar**, cochichar, **conchavar**, **consolar**, **debicar**, **debochar**, **desiludir**, diagnosticar, **escarnecer**, **ferroar**, **inclinarse**, **interceptar**, **maldizer**, participar, **remediar**, **suspirar**, sussurrar, **zombar**. Os verbos em negrito pertencem todos a uma mesma categoria de Moura Neves (2000), *verbos que instrumentalizam ou circunstanciam o que é dito*, que são os verbos que podem introduzir discurso, mas que não necessariamente indicam atos de fala. Diante desse número, e também porque, apesar do grande número de verbos encontrados, sabemos que a nossa lista não é exaustiva, resolvemos, então, testar a lista de Moura Neves (2000) nos três corpora monolíngues usados neste estudo.

Primeiramente, queríamos entender por que esses 25 verbos não apareceram entre os 293 verbos que compilamos. Consultamos então todo o material do AC/DC e identificamos três motivos diferentes para a ausência: ou (i) tinham apenas uma ocorrência (e, portanto, foram descartados na nossa seleção inicial), ou (ii) os verbos não foram encontrados entre as ocorrências, ou (iii) as ocorrências não correspondiam a VE (exemplo: “Com o vídeo ‘Circular’, *participou* este ano do prestigiado ‘Curta Cinema’, Festival Internacional de Curtas do Rio”).

Em seguida, precisávamos confirmar se tais verbos poderiam funcionar como VE (e, em caso afirmativo, seriam incluídos em nossa lista). Dos 25 verbos, quinze foram empregados como VE em todos<sup>9</sup> os corpora.

O fato de não termos encontrado dez dos verbos propostos em Moura Neves (2000) não significa necessariamente que tais verbos não possam ser empregados como VE. Devemos lembrar que, apesar da dimensão dos corpora consultados, um corpus jamais será exaustivo e tampouco dará conta de todos os fenômenos da língua. Como o corpus usado por Moura Neves (2000) não está disponível para o público, não temos como verificar se as ocorrências analisadas na elaboração da gramática correspondem, para nós, a VE. Ainda assim, vale ressaltar que a ausência desses dez verbos em corpora que, juntos, somam mais de um bilhão de palavras, pode indicar que o emprego dos verbos como introdutores de discurso relatado seja incomum.

Além disso, traçamos também um paralelo entre o estudo de Moura Neves (2000) e esta pesquisa no que se refere ao tipo de discurso. Um dos trabalhos de Moura Neves (2000) consistiu em determinar se os VE são aceitos no discurso direto (DD) e/ou no indireto (DI). Reproduzimos a classificação da autora, considerando (a) os verbos em que houve divergência quanto ao tipo de discurso associado ao verbo, ou seja, quando nossos resultados e os de Moura Neves (2000) diferem; e (b) os verbos encontrados apenas por nós – contribuindo, assim, com o levantamento já feito pela autora. Por questões de espaço, não apresentamos o quadro aqui, mas ele está disponível em (FREITAS, 2016).

Diante da análise feita acima, merecem destaque os cinco verbos que não introduziram citação direta em nenhuma ocorrência: *acrescer*, *debater*, *discutir*, *presumir* e *proibir*. Novamente, ressaltamos que o fato de não encontrarmos ocorrências desses verbos no discurso direto não significa, necessariamente, que o emprego nesse tipo de contexto não seja possível. Como já afirmamos, apesar de os corpora analisados apenas darem conta de apenas uma parcela do que pode ser produzido na língua portuguesa, a não ocorrência pode, talvez, indicar que o emprego desses verbos como VE seja pouco comum.

#### 4 CONSIDERAÇÕES FINAIS

Neste trabalho, apresentamos a metodologia para a elaboração de um léxico dos verbos de elocução em português (VE). Trata-se de uma parcela de um trabalho mais amplo, que engloba também a elaboração do DISSE, um glossário de VE que tem como principal objetivo auxiliar profissionais de tradução (FREITAS, 2016). Para tanto, fizemos um amplo estudo descritivo dessa classe de verbos.

<sup>9</sup> Lembramos que a utilização de todos os corpora corresponde a uma busca em mais de um bilhão de palavras, distribuídas em diferentes gêneros textuais.

O trabalho tomou por base grandes corpora da língua portuguesa e dialogou ainda com a Linguística Computacional, quando toma a tarefa de identificação de citação também como motivação. De fato, os verbos e os padrões associados que foram descritos aqui já foram incorporados na elaboração de um sistema de detecção automática de citações em português.

Também resultado do léxico e dos padrões, está em andamento a anotação dos corpora do projeto AC/DC, permitindo, assim, a busca pelo campo semântico *dizer*, processo indicado em Freitas (2016). Como explicitamos ao longo deste artigo, o material é público e está disponível para consultas on-line.

Com base em corpus, estabelecemos oito padrões gerais relativos aos VE, a partir dos quais foi possível um levantamento desses verbos e, assim, construímos um grande léxico dos VE do português, constituído pelo número impressionante de 308 verbos (os 293 verbos encontrados com a nossa abordagem somados aos os quinze verbos encontrados apenas por Moura Neves (2000), porém confirmados por nós), todos validados manualmente com ocorrências de uso.

Em termos quantitativos, o fato de termos compilado 293 verbos, isto é, mais verbos do que qualquer estudo conhecido por nós, seja em português, seja em outras línguas, aponta o sucesso do caminho que escolhemos – o caminho da consulta em grandes corpora. A apreensão de padrões envolvidos no relato, isto é, a observação de que um mesmo tipo de citação (direta, indireta ou mista) pode ser expressa por meio de diferentes padrões também são contribuições que julgamos valiosas, e que explicitam as potencialidades dos estudos descritivos com base em grandes corpora da nossa língua.

Por fim, esperamos ter demonstrado, ao longo deste trabalho, como pode ser produtivo o diálogo entre a descrição de uma língua, a tradução e a linguística computacional. A descrição, de qualquer tipo, não pode ser uma atividade abstrata, descolada de intenções e objetivos, o que leva a nossa pesquisa a apontar para a adequação de parcerias com a Tradução e com a Linguística Computacional, áreas aplicadas.

## REFERÊNCIAS

AUSTIN, J. L. *How to do things with words*. 2. ed. Cambridge: Harvard University Press, 1962.

BIBER, D. et al. *Longman grammar of spoken and written English*. Harlow: Pearson Education ESL, 1999.

COSTA, L.; SANTOS, D.; ROCHA, P. A. Estudando o português tal como é usado: o serviço AC/DC. In: PARDO, T., NUNES, M.G.V. *The 7th Brazilian symposium in information and human language technology (STIL 2009)*, São Carlos, [2009]. Disponível em: <[http://nilc.icmc.usp.br/til/stil2009\\_English/Proceedings/stil/Costa-57572\\_1.pdf](http://nilc.icmc.usp.br/til/stil2009_English/Proceedings/stil/Costa-57572_1.pdf)>. Acesso em: 18 ago 2017.

FRANKENBERG-GARCIA, A; SANTOS, D. COMPARA, um corpus paralelo de português e de inglês na Web. *Cadernos de Tradução*, Florianópolis, v.1, n. 9, p. 61-79, 2002.

FREITAS, B. *O dizer em português: diálogos entre tradução, descrição e linguística computacional*. 2016. 116f. Dissertação (Mestrado em Estudos da Linguagem) – Programa de Pós-Graduação em Estudos da Linguagem, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

FREITAS, C.; ROCHA, P.; BICK, E. Floresta sintá(c)tica: bigger, thicker and easier. In: TEIXEIRA, A. et al. (Ed.). *Computational processing of the portuguese language: 8th international Conference, PROPOR 2008*. Germany: Springer Verlag, 2008. p. 216-219.

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

HOUAISS, A.; VILLAR, M. S. *Grande dicionário Houaiss*. Verbetes “Considerar”, [2009]. Disponível em: <<https://houaiss.uol.com.br/>>. Acesso em: 25 ago. 2016.

KIPPER, K. et al. Extending VerbNet with novel verb classes. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC 2006), 5., *Proceedings...* Genova, Itália, 2006.

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

MARTINS, E. *Manual de redação e estilo do Estado*. Verbetes “Verbos mais que errados”. Disponível em: <<http://www.estadao.com.br/manualredacao/>>. Acesso em: 10 jan. 2016.

MOURA NEVES, M. H de. *Gramática de usos do português*. São Paulo: Editora Unesp, 2000.

PARETI, S.; O'KEEFE, T.; KONSTAS, I.; CURRAN, J. R. e KOPRINSKA, I. Automatically detecting and attributing indirect quotations. In: YAROWSKY, D. et al. (Ed.). *Proceedings of the 2103 conference on empirical methods in natural language processing (EMNLP 2013)*, Washington, USA, 2013. p. 989-999. Disponível em: <<http://www.aclweb.org/anthology/D/D13/D13-1101.pdf>>. Acesso em: 18 ago. 2017.

SAGOT, B.; DANLOS, L.; STERN, R. A lexicon of French quotation verbs for automatic quotation extraction. In: CALZOLARI, N. et al. (Ed.). *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*, La Valette, Malta, 2010. p. 294-299. Disponível em: <[http://www.lrec-conf.org/proceedings/lrec2010/pdf/387\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/387_Paper.pdf)>. Acesso em: 18 ago. 2017.

SANTOS, D. Gramateca: corpus-based grammar of Portuguese. In: BAPTISTA, J. et al. (Ed.). *Computational processing of Portuguese: 11th international conference (PROPOR 2014)*. Germany: Springer, 2014. p. 214-219.

SANTOS, D.; ROCHA, P. The key to the first CLEF in Portuguese: Topics, questions and answers in CHAVE. In: PETERS, C. et al. (Ed.). *Multilingual information access for text, speech and images*, Workshop of the Cross-Language Evaluation Forum, 5, Bath, 2004, Revised Selected Papers. Berlin/Heidelberg: Springer, Lecture Notes in Computer Science, 2005. p. 821-832.

SCARTON, C.; ALUISIO, S. Towards a cross-linguistic VerbNet-style lexicon to Brazilian Portuguese. In: LAMBERT, P. et al. (Ed.). *Proceedings of the LREC 2012 workshop on creating cross-language resources for disconnected languages and styles (CREDISLAS 2012)*, Istanbul, 2012. p. 11-18. Disponível em: <<http://www.lrec-conf.org/proceedings/lrec2012/index.html>>. Acesso em: 15 ago. 2017.

TRUGGO, L. F. *Classes de palavras - da Grécia Antiga ao Google: um estudo motivado pela conversão de tagsets*. 2016. 113f. Dissertação (Mestrado em Letras/Estudos da Linguagem) – Programa de Pós-Graduação em Estudos da Linguagem, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Recebido em 24/10/2016. Aceito em 17/12/2016.