

A. MESAS REDONDAS

MÉTODOS DE INVESTIGAÇÃO EM CORPORA: FERRAMENTAS PARA CLASSIFICAÇÃO DE DADOS EXTRAÍDOS DE CORPORA DE PEQUENAS DIMENSÕES PARA ANÁLISES DISCURSIVAS*

Daniel Alves¹
Roberto Carlos de Assis²*Resumo*

A Linguística de Corpus ocupa, atualmente, papel de destaque entre as investigações desenvolvidas no campo disciplinar dos Estudos da Tradução. No artigo aqui apresentado, pretendemos apresentar e discutir dois métodos de investigação em corpora, aplicáveis a pesquisas desenvolvidas a partir de corpora de pequenas dimensões e baseadas na classificação de dados. O primeiro desses métodos propõe a anotação de corpora por meio de categorias numéricas e o segundo propõe o uso de planilhas eletrônicas para o processamento, organização e análise de dados. Dentre as vantagens do primeiro método estão a forma de anotação no próprio corpus e a agilidade de digitação e dentre as vantagens do segundo método estão a possibilidade de utilização dos recursos de planilhas, especialmente no que se refere a organização e contagem de dados, sistematização de números e elaboração de gráficos.

Palavras-chave: Linguística de Corpus; Métodos de Pesquisa; Classificação de dados linguísticos

Abstract

Corpus Linguistics has played a prominent role among investigations within Translation Studies. In this paper we intend to present and discuss two investigation methods which can be applied to small corpus investigation involving classification of data. The first one proposes a tagging system using numerical categories while the second proposes the use of electronic spreadsheets for processing, organizing and data analysis. Advantages of the first method include in-line annotation and typing agility, while the second adds the possibilities of using spreadsheet resources to organize and count data, systematize numbers and elaborate graphs.

Key-words: Corpus Linguistics; Methods of investigation; Classification of linguistic data

1. Introdução

A Linguística de Corpus ocupa, atualmente, papel de destaque entre as investigações desenvolvidas no campo disciplinar dos Estudos da Tradução. Ao analisar teses e dissertações brasileiras, Alves e Vasconcellos (no prelo) apontam a Linguística de Corpus como ferramenta empregada por 17,5% dos trabalhos que investigam a tradução entre os anos de 2006 e 2010 — um percentual elevado, considerando-se o caráter nômade e a diversidade de afiliações de trabalhos acadêmicos abraçados pelos Estudos da Tradução. Analisando os percursos metodológicos mais frequentemente empregados nas explorações de corpora de pequenas dimensões, os autores identificam uma proeminência de procedimentos de anotação³ como método adotado por pesquisadores(as) da área para se produzir um levantamento estatístico, visando o acompanhamento de padrões discursivos e decisões tradutórias, entre outros aspectos textuais.

* Gostaríamos de agradecer à Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo financiamento dos anais da VII Escola Brasileira de Linguística Computacional e do XIII Encontro de Linguística de Corpus, processo nº 3472/2015-87.

¹ Daniel Antonio de Sousa Alves (daniel@cchla.ufpb.br) é Professor da Universidade Federal da Paraíba (UFPB). Seus interesses de pesquisa incluem Linguística de Corpus, metodologia de trabalho científico, Linguística Sistemico-Funcional, Tradução e Conflito.

² Roberto Carlos de Assis (robertoassis@cchla.ufpb.br) é Professor da Universidade Federal da Paraíba (UFPB). Seus interesses de pesquisa incluem Estudos da Tradução, Linguística de Corpus e Representação de Atores Sociais.

³ Para Hunston (2002) anotação é o processo de acrescentar informações a um corpus com o objetivo de interpretá-lo linguisticamente

Tendo um corpus e um objetivo em mãos, um dos pontos de partida de uma investigação é a definição de categorias a serem utilizadas para anotação e posterior interpretação. Leech (2005) lista seis tipos de anotações em corpora⁴: anotações fonéticas, semânticas, pragmáticas, discursivas, estilísticas e lexicais. Embora pesquisas possam ser feitas sem estas anotações no corpus, ou existam *softwares* de anotação automática, corpora de pequenas dimensões proporcionam a possibilidade de desenvolvimento de pesquisas com categorias que, nem sempre, são acomodadas pelos *softwares* existentes. O que anotar dependerá dos interesses do pesquisador e as possibilidades de constituição de categorias são amplas. Geralmente advêm de teorias ou sistemas preexistentes como aquelas das diversas gramáticas, de sistemas como o de representação de atores sociais (VAN LEEUWEN, 1996) ou de Avaliatividade (MARTIN E WHITE, 2005), de listas de procedimentos técnicos ou estratégias de tradução, entre outras possibilidades. Ressalve-se que podem advir, também, de categorias desenvolvidas pelo próprio pesquisador.

Neste artigo discutiremos duas formas de classificação de dados em corpora (alternativas aos *softwares* de anotação automática ou semiautomática — ver Lima-Lopes, neste volume), a saber: a anotação por meio de etiquetas numéricas e o uso de planilhas eletrônicas. Uma ressalva a ser feita, já de antemão, a ambos os métodos é sua abrangência a corpora de dimensões que não sejam nem tão pequenos a ponto de não necessitarem de ferramentas computacionais para levantamento estatístico, nem tão grandes a ponto de uma intervenção humana antecipada — mencionada por Sinclair (2001) em sua definição de corpora de pequenas dimensões — ser inviável.

Este trabalho está desenvolvido em três seções, além desta introdução, organizadas da seguinte maneira. A segunda seção, a seguir, apresenta a proposta de anotação manual de corpora utilizando etiquetas numéricas. A terceira seção apresenta a proposta de classificação de dados extraídos de corpora utilizando planilhas eletrônicas, mostrando formas de uso e indicando o uso de ferramentas de contagem condicional. A quarta e última seção apresenta as conclusões, fazendo um breve apanhado do trabalho aqui apresentado.

2. Rotulação de corpora com base em códigos numéricos

Uma alternativa aos programas automáticos de anotação de corpora é a anotação manual através do desenvolvimento de sua própria grade de anotação, ou código de rotulação. Estas podem ser inseridas no próprio corpus entre parênteses angulares < >, para não interferir nos dados quantitativos, ou através de planilha eletrônica. Esta seção discutirá a primeira alternativa.

A inserção da anotação no corpus pode ser através do próprio nome das categorias ou uma abreviação, com ou sem espaços, como no exemplo 01.

Exemplo 01: ... resmungou uma voz <HOD2002> <EUROPEU> <PROCESSO VERBAL> <DIZENTE>, e eu <HOD2002> <EUROPEU> <PROCESSO MENTAL> <EXPERIENCIADOR> soube então que havia pelo menos mais um ouvinte acordado <HOD2002> <EUROPEU> <PROCESSO EXISTENCIAL> <EXISTENTE> além de mim.
Fonte: O Coração das trevas (2002)

O exemplo 01 ilustra a inserção de várias anotações cocorrentes. Identificam-se o subcorpus <HOD2002>, o grupo de atores sociais <EUROPEU>, o tipo de Processo⁵ <PROCESSO VERBAL> e o tipo de participante associado ao Processo <DIZENTE>. Neste caso, o pesquisador, semelhante a Assis (2009), estaria interessado nas formas de representação de europeus e de africanos em *Heart of darkness* e em duas de suas traduções (1984 e 2002) e teria utilizado uma grade como a apresentada no QUADRO 1:

⁴ Acrescentem-se anotações morfossintáticas automáticas do tipo *POS tagging* (etiquetamento de partes do discurso)

⁵ Termo da Gramática Sistemico-Funcional

Quadro 1 — Exemplo de grade de marcação

Subcorpora	Grupos de atores sociais	Tipos de Processos	Tipos de Participantes
Heart of Darkness (HOD)	Europeus	Material	Ator
Tradução 1984 (HOD1984)	Africanos	Mental	Meta
Tradução 2002 – HOD (2002)		Existencial	Extensão
		Verbal	Experienciador
		...	Fenômeno
			Existente
			Dizente
			Verbiagem
			...

Uma crítica a este método, apontada por Feitosa (2005), é que o uso de palavras às vezes extensas, mesmo com abreviaturas, pode causar erros de digitação, comprometendo o levantamento dos dados através de um concordanciador.

Uma alternativa a esta forma de anotação seria a utilização de códigos numéricos ao invés dos nomes das categorias. Feitosa (2005), por exemplo, desenvolveu o CROSF-14, Código de Rotulação Sistemico-Funcional, para anotação e análise de tipos de Tema, categoria da Linguística Sistemico-Funcional associada à metafunção textual. O autor ressalta a funcionalidade e praticidade do uso de algarismos em vez dos próprios nomes das categorias. Segundo o autor, a concentração destes ao lado direito de certos teclados permite a utilização de uma única mão para digitação ao mesmo tempo em que reduz a quantidade de toques no teclado, limitando, assim, a probabilidade de erros, entre outras vantagens.

O código de rotulação pode ser desenvolvido com quaisquer categorias, além daquelas de Feitosa. Tomando o exemplo acima, poderíamos desenvolver uma grade de anotação ou código de rotulação como a apresentada no QUADRO 2, em que a posição *a* indica os subcorpora; a posição *b*, os grupos de atores sociais; a posição *c*, os tipos de processos; e a posição *d*, os tipos de participantes. Na posição *c* poderíamos elencar os demais tipos de processos, assim como na posição *d*, os demais tipos de participantes (indicados pelas reticências). Note a repetição dos algarismos na posição *d*, que é dependente da marcação da posição anterior. Os tipos de participantes (posição *d*) estão diretamente associados aos tipos de processos (posição *c*).

Quadro 2 — Exemplo de grade de marcação com código numérico

a Subcorpora	b Grupos de atores sociais	c Tipos de Processos	d Tipos de Participantes	
1 Heart of Darkness	1 Europeus	1 Material	Pr.	1 Ator
2 Tradução 1984	2 Africanos	2 Mental	Material	2 Meta

3 Tradução 2002		3 Existencial		3 Extensão
		4 Verbal	Pr. Mental	1 Experienciador
		...		2 Fenômeno
			Pr. Existencial	1 Existente
			Pr. Verbal	1 Dizente
				2 Verbiagem
		

Ao aplicarmos a grade de marcação do QUADRO 2 no exemplo anterior, teríamos a seguinte configuração:

... resmungou uma voz <3141>, e eu <3121> soube então que havia pelo menos mais um ouvinte acordado < 3131> além de mim. (Fonte: O Coração das trevas (2002))

Embora no início possa parecer mais complicada que a anotação pelo nome das categorias, o avançar do trabalho familiariza o pesquisador com os códigos e a marcação torna-se mais ágil, além de deixar o corpus menos poluído. Outra vantagem deste tipo de marcação é que ela pode ser feita no próprio corpus, em formato *.doc*, e posteriormente convertida para o formato *.txt* para exploração através de um concordanciador.

Para levantamento dos dados quantitativos, utilizando-se o *AntConc*, por exemplo, utilizam-se as etiquetas como nóculo de busca. Tomando como exemplo a grade apresentada no Quadro 2, caso o interesse do pesquisador seja no número de ocorrências em que o os europeus são representados como Dizente no original, deve-se utilizar como nóculo de busca <1241>. Para o levantamento de ocorrências nas diversas subcategorias, recorre-se ao sinal de interrogação (?) como dígito coringa⁶. Por exemplo, caso o interesse recaia sobre o número de ocorrências de referências aos africanos na tradução de 1984, utiliza-se o nóculo <22??>. Neste caso, as linhas de concordância apresentadas conterão ocorrências de referências aos africanos na tradução de 1984, independente do tipo de processo em que estão inscritos ou do tipo de participação. A busca pode ser refinada, substituindo-se as interrogações pelo código de interesse. Por exemplo, caso o interesse seja apenas as referências aos africanos nas quais eles estão inscritos em Processos Mentais, o nóculo de busca deve ser <222?>. Neste caso, as linhas de concordância apresentariam apenas as ocorrências deste tipo de Processo, incluindo-se as participações como Experienciador ou Fenômeno, que são os Participantes inscritos neste tipo de Processo. Comparações podem ser feitas com o original ou com a tradução de 2002 utilizando-se os nósculos de busca <122?> e <322?>, respectivamente.

Um outro exemplo de utilização de grade de marcação que segue um código numérico é Novais (2015). A autora estava interessada nas formas de representação dos manifestantes e dos representantes do governo em jornais nacionais e internacionais por ocasião das manifestações de junho de 2013 no Brasil. Para tanto desenvolveu a grade de marcação apresentada no QUADRO 3.

Quadro 3 — Grade de marcação de Novais (2015)

A Corpus	B	C	D Determ./	E	F
-------------	---	---	---------------	---	---

⁶ Em alguns navegadores o dígito coringa é o asterisco (*)

	Inclusão / Exclusão	Person. / Imperson.	Indeterm.	Forma de Pers./ Impers.	Grupo de Atores
1 Correio Braziliense	1 Inclusão	1 Personalização	1 Determinação	1 Nomeação	1 Manifestantes
2 Folha de São Paulo	2 Exclusão	2 Impersonalização	2 Indeterminação	2 Funcionalização	2 Representantes do Governo
3 Jornal do Brasil		0 Não se aplica	0 Não se aplica	3 Classificação	
4 Chicago Tribune				4 Id. Relacional	
5 New York Times				5 Id. Física	
6 The Guardian				6 Id. p/ vestuário	
				...	

Fonte: Adaptado de Novais (2015, p. 48)

O QUADRO 3 apresenta um maior número de categorias do que o QUADRO 2, indicando que o número de posições na grade dependerá dos interesses de pesquisa. Vale destacar, também, a utilização do código zero (0) para os casos em que a classificação de uma determinada ocorrência não se aplique. Neste exemplo, o código zero foi utilizado para as ocorrências de exclusão dos atores sociais, em que se tem o código 2 na posição *b* (não se fala de Personalização ou Impersonalização em casos de Exclusão). Seguindo a grade de marcação do QUADRO 3, lê-se as marcações <111131> e <120001> no exemplo 02, a seguir, da seguinte forma:

Exemplo 02: Quase 2 milhões de brasileiros <111131> fizeram manifestações <120001> pela redução das passagens do transporte público, contra os gastos com as obras da Copa do Mundo, pelo aumento dos recursos para a saúde e educação e contra a corrupção e a impunidade. (Fonte: Novais (2015, p. 49).

A marcação <111131> indica que esta é uma representação dos manifestantes (posição *f*); ocorre no *Correio Braziliense* (posição *a*); é um caso de inclusão (posição *b*) por personalização (posição *c*). O ator social é determinado (posição *d*) e classificado (posição *e*). Já a marcação <120001> indica uma exclusão (posição *b*) dos manifestantes (posição *f*) no subcorpus *Correio Braziliense* (posição *a*).

Vale destacar que a construção da grade de marcação pode sofrer modificações ao longo da pesquisa, caso verifique-se a necessidade de revisão dos objetivos. Não são raros os casos em que as ocorrências no corpus nos fazem refletir sobre e reavaliarmos as categorias inicialmente propostas para análise. É o caso de Assis (2009), que chegou a sete versões de sua grade de marcação ou Feitosa (2005), que chegou à versão catorze de seu protocolo. Para evitar retrabalho, aconselha-se a testar a viabilidade da grade em uma pequena porção do corpus antes de partir para a marcação definitiva.

A principal desvantagem deste tipo de marcação é que, embora as buscas com as linhas de concordância ofereçam a quantidade exata de ocorrências de cada nóculo de busca, as tabelas e gráficos a serem apresentados devem ser confeccionados através de outras ferramentas baseadas em anotações manuais.

A próxima seção apresenta o uso de planilhas eletrônicas como alternativa para a classificação de dados levantados a partir de corpora.

3. Uso de planilhas eletrônicas na classificação e análise de corpora

Com o objetivo de mostrar funcionalidades de planilhas eletrônicas aplicáveis à pesquisa (baseada em classificação de dados) com corpora, esta seção investiga uma amostra do corpus monolíngue composto pelo texto traduzido *Mas não se mata cavalo?*⁷ de 1947 (tradução de Érico Veríssimo do romance *They shoot horses, don't they?*, de 1935, por Horace McCoy). Na investigação, são analisados os usos do pronome 'eu' no corpus a partir do referencial teórico de Maia (1998). O estudo utiliza o concordanciador *AntConc 3.4.3w* (ANTHONY, 2011), por se tratar de uma opção gratuita, que não exige instalação e que disponibiliza um volume considerável de recursos para investigação de corpora.

Por se tratar de uma indicação metodológica, exemplificada por meio do estudo de uma amostra dos dados levantados, esta seção mescla elementos comuns a seções de 'Revisão teórica' e 'Método de pesquisa'. Nos parágrafos que se seguem, são apresentados brevemente o trabalho de Maia (1998) que oferece o aporte teórico para esta discussão, os métodos de uso do software *AntConc 3.4.3w* para o levantamento de dados, as categorias de análise aqui empregadas e os métodos utilizados para classificar os dados e contabilizá-los utilizando planilhas eletrônicas.

A escolha pelo trabalho de Maia (1998) para a realização deste estudo se justifica pela facilidade em replicar a categorização proposta pela autora, sem a necessidade de estabelecer extensas discussões teóricas (considerando as limitações de espaço para a investigação aqui apresentada). Em seu artigo de 1998, Maia discute os padrões de estruturas oracionais e as funções sujeito, tema e tópico e investiga os usos dos pronomes de primeira pessoa ('eu' e 'I') em corpora ficcionais, levando em consideração questões como posição (anteposta ou posposta) do pronome; realização explícita e elisão do pronome; e recuperação do sujeito a partir de desinências verbais.

A investigação apresentada por Maia (1998) mostra indícios de que, em textos traduzidos (de inglês) para português, a proporção de uso de pronomes é maior do que a observada em textos originalmente escritos em português — o que a autora atribui a uma influência do texto fonte sobre o texto traduzido, após fazer considerações sobre a recuperação de sujeitos oracionais a partir das desinências verbais. Além disso, Maia (1998, p.9) identifica um número considerável de pronomes em posição pós-verbal — o que chama a atenção da autora, considerando que, “no ensino de português como segunda língua, aprendizes são desencorajados(as) a utilizar [posposições] de pronome”.

O primeiro passo para o levantamento dos dados a serem aqui analisados é, no software *AntConc 3.4.3w*, utilizar a caixa para levantamento de linhas de concordância na aba 'Concordance'. A Figura 1, a seguir, mostra o levantamento de linhas de concordância no corpus, utilizando 'eu' como nóculo de busca:

⁷ Nossos agradecimentos ao Grupo CORDIAL (<http://letra.letras.ufmg.br/cordial/>), da Universidade Federal de Minas Gerais, pelo corpus.

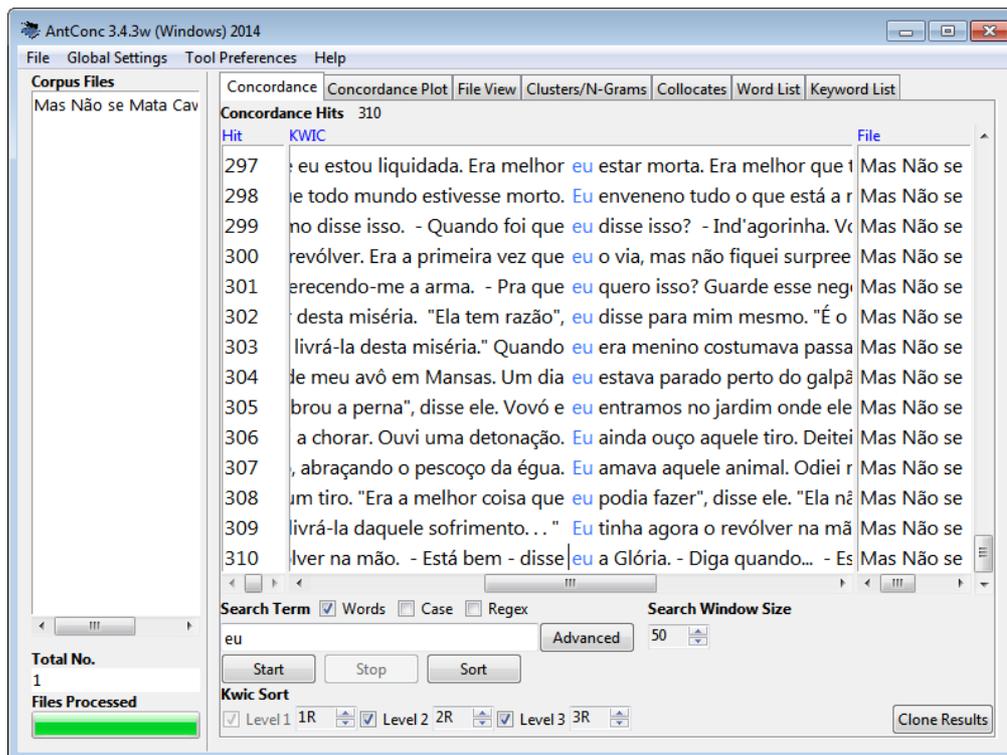


Figura 1 — Levantamento das linhas de concordância a partir do nóculo 'eu'

A Figura 1, acima, mostra parte das 310 linhas de concordância geradas pelo software a partir do levantamento do nóculo 'eu' no corpus investigado. Essas linhas podem ser copiadas e organizadas em planilhas eletrônicas ou, utilizando-se a função 'Save Output to Text File', salvas em um arquivo TXT. As Figuras 2 e 3, a seguir, mostram o uso da função 'Save Output to Text File' (aqui indicada por facilitar o processo de organização dos dados em planilhas eletrônicas).

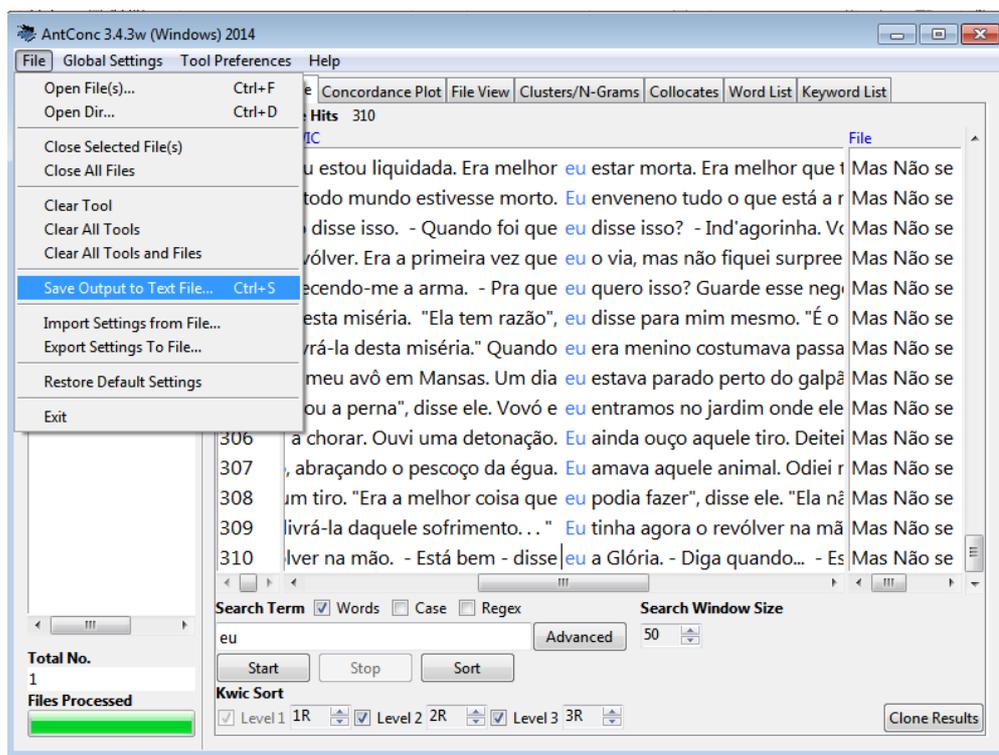
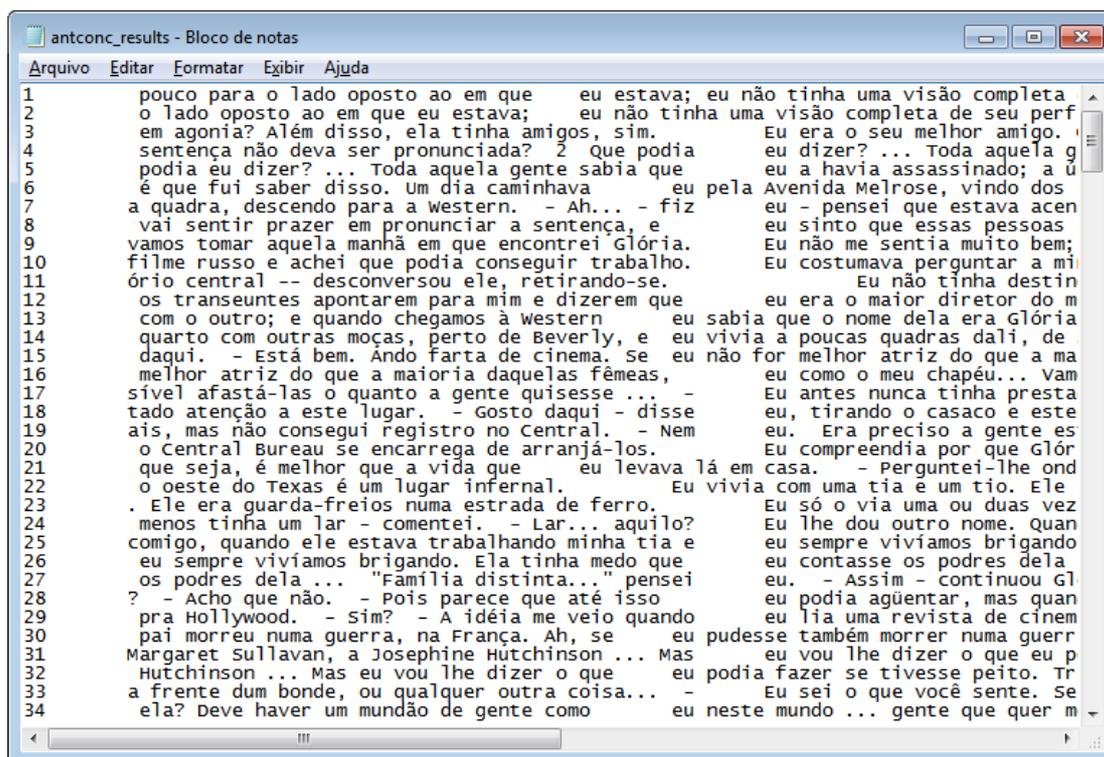


Figura 2 — Salvando o resultado do processo de levantamento em um arquivo externo (tipo TXT)

A Figura 2, acima, mostra o uso da opção 'Save Output to Text File' no software AntConc 3.4.3w. Como anteriormente anunciado, a função gera um arquivo TXT, mostrado na Figura 3, a seguir:



```
1 pouco para o lado oposto ao em que eu estava; eu não tinha uma visão completa
2 o lado oposto ao em que eu estava; eu não tinha uma visão completa de seu perf
3 em agonia? Além disso, ela tinha amigos, sim. Eu era o seu melhor amigo.
4 sentença não deva ser pronunciada? 2 que podia eu dizer? ... Toda aquela g
5 podia eu dizer? ... Toda aquela gente sabia que eu a havia assassinado; a u
6 é que fui saber disso. Um dia caminhava eu pela Avenida Melrose, vindo dos
7 a quadra, descendo para a western. - Ah... - fiz eu - pensei que estava acen
8 vai sentir prazer em pronunciar a sentença, e eu sinto que essas pessoas
9 vamos tomar aquela manhã em que encontrei Glória. Eu não me sentia muito bem;
10 filme russo e achei que podia conseguir trabalho. Eu costumava perguntar a mi
11 ório central -- desconversou ele, retirando-se. Eu não tinha destin
12 os transeuntes apontarem para mim e dizerem que eu era o maior diretor do m
13 com o outro; e quando chegamos à western eu sabia que o nome dela era Glória
14 quarto com outras moças, perto de Beverly, e eu vivia a poucas quadras dali, de
15 daqui. - Está bem. Ando farta de cinema. Se eu não for melhor atriz do que a ma
16 melhor atriz do que a maioria daquelas fêmeas, eu como o meu chapéu... Vam
17 sível afastá-las o quanto a gente quisesse ... - Eu antes nunca tinha presta
18 tado atenção a este lugar. - Gosto daqui - disse eu, tirando o casaco e este
19 ais, mas não consegui registro no Central. - Nem eu. Era preciso a gente es
20 o Central Bureau se encarrega de arranjá-los. Eu compreendia por que Glór
21 que seja, é melhor que a vida que eu levava lá em casa. - Perguntei-lhe ond
22 o oeste do Texas é um lugar infernal. Eu vivia com uma tia e um tio. Ele
23 . Ele era guarda-freios numa estrada de ferro. Eu só o via uma ou duas vez
24 menos tinha um lar - comentei. - Lar... aquilo? Eu lhe dou outro nome. Quan
25 comigo, quando ele estava trabalhando minha tia e eu sempre vivíamos brigando
26 eu sempre vivíamos brigando. Ela tinha medo que eu contasse os podres dela
27 os podres dela ... "Família distinta..." pensei eu. - Assim - continuou Gl
28 ? - Acho que não. - Pois parece que até isso eu podia agüentar, mas quan
29 pra Hollywood. - sim? - A idéia me veio quando eu lia uma revista de cinem
30 pai morreu numa guerra, na França. Ah, se eu pudesse também morrer numa guerr
31 Margaret Sullavan, a Josephine Hutchinson ... Mas eu vou lhe dizer o que eu p
32 Hutchinson ... Mas eu vou lhe dizer o que eu podia fazer se tivesse peito. Tr
33 a frente dum bonde, ou qualquer outra coisa... - Eu sei o que você sente. Se
34 ela? Deve haver um mundão de gente como eu neste mundo ... gente que quer m
```

Figura 3 — Arquivo TXT resultante do salvamento mostrado na Figura 2 (anterior)

A Figura 3 mostra o arquivo TXT gerado a partir do software *AntConc 3.4.3w*. No arquivo, estão organizados os resultados da busca apresentada na Figura 1 (desta seção). Uma vez salvos, esses resultados dessa busca podem ser classificados utilizando procedimentos de etiquetamento e rotulação de dados (como o descrito na seção 2 deste texto), ou copiados e colados em planilhas eletrônicas. Nesta seção, como anteriormente definido, são utilizadas planilhas eletrônicas.

Dentre as vantagens dessa opção, estão, como aponta Alves (2014), a acessibilidade dessas ferramentas (atualmente disponíveis em grande parte dos computadores), a versatilidade e as funções oferecidas por tais softwares — em termos de “organização, contagem (e contagem condicional) de dados, sistematização de números e elaboração de gráficos, por exemplo”, como aponta Alves (2014, p.120) — além da portabilidade dessas ferramentas e da compatibilidade dos seus resultados com outros *softwares* (como processadores de textos, por exemplo).

Para a investigação aqui mostrada, o conteúdo do arquivo TXT (mostrado na Figura 3) é copiado do bloco de notas e colado em uma planilha eletrônica. O resultado desse processo é mostrado na Figura 4, a seguir:

A	B	C	D	E	F	G
1	Hit	KWIC				
2	1 pouco para o lado oposto ao em que	eu estava; eu não tinha uma visão completa de				
3	2 o lado oposto ao em que eu estava;	eu não tinha uma visão completa de seu perfil,				
4	3 em agonia? Além disso, ela tinha amigos, sim.	eu era o seu melhor amigo. O seu único				
5	4 sentença não deva ser pronunciada? 2 Que podia	eu dizer? ... Toda aquela gente sabia que eu a				
6	5 podia eu dizer? ... Toda aquela gente sabia que	eu a havia assassinado; a única pessoa que me				
7	6 é que fui saber disso. Um dia caminhava	eu pela Avenida Melrose, vindo dos estúdios da Par				
8	7 a quadra, descendo para a Western. - Ah... - fiz	eu - pensei que estava acenando pra mim. - A troc				
9	8 vai sentir prazer em pronunciar a sentença, e	eu sinto que essas pessoas atrás de mim vão				
10	9 vamos tomar aquela manhã em que encontrei Glória.	eu não me sentia muito bem; estava ainda um				
11	10 filme russo e achei que podia conseguir trabalho.	eu costumava perguntar a mim mesmo se haveria cois				
12	11 ório central -- desconversou ele, retirando-se.	eu não tinha destino certo; estava apenas dirigind				
13	12 os transeuntes apontarem para mim e dizerem que	eu era o maior diretor do mundo... Foi nesse				
14	13 com o outro; e quando chegamos à Western	eu sabia que o nome dela era Glória Beatty,				
15	14 quarto com outras moças, perto de Beverly, e	eu vivia a poucas quadras dali, de sorte que				
16	15 daqui. - Está bem. Ando farta de cinema. Se	eu não for melhor atriz do que a maioria				
17	16 melhor atriz do que a maioria daquelas fêmeas,	eu como o meu chapéu... Vamos nos sentar no				
18	17 sível afastá-las o quanto a gente quisesse ... -	eu antes nunca tinha prestado atenção a este lugar				
19	18 tado atenção a este lugar. - Gosto daqui - disse	eu , tirando o casaco e estendendo-o sobre a				
20	19 ais, mas não consegui registro no Central. - Nem	eu . Era preciso a gente estar registrado no Centr				
21	20 o Central Bureau se encarrega de arranjá-los.	eu compreendia por que Glória não conseguira regis				
22	21 que seja, é melhor que a vida que	eu levava lá em casa. - Perguntei-lhe onde vivia				
23	22 o oeste do Texas é um lugar infernal.	eu vivia com uma tia e um tio. Ele				

Figura 4 — Organização dos dados (do arquivo TXT mostrado na Figura 3) em planilha eletrônica

Como anteriormente anunciado, a Figura 4, acima, mostra as linhas de concordância resultantes do levantamento aqui realizado organizadas em uma planilha eletrônica. A numeração de cada linha (mostrada na Coluna A) e a divisão das linhas de concordância (mostrada entre as colunas B e C) são resultado dos processos de copiar e colar os dados gerados pelo concordanciador. Cumpre observar que as intervenções feitas nos dados (mostrados na Figura 4) são: colorir o pronome ‘eu’ (para indicar o nódulo de cada linha) e excluir informações adicionais (nome de arquivo, por exemplo) para facilitar a visualização dos dados.

Realizada a organização dos dados na planilha, passa-se à delimitação e classificação dos dados. A delimitação tem por objetivo selecionar para análise apenas os casos em que o pronome ‘eu’: 1) constitui sujeito oracional das linhas de concordância; e 2) constrói ideia de sujeito singular (não sendo parte constituinte de um sujeito composto). A classificação dos dados segue as categorias investigadas por Maia (1998) quanto à posição pronominal e quanto ao papel na ênfase do pronome enquanto enfatizador ou definidor do sujeito. A seguir, são apresentados dois exemplos que ilustram essas classificações:

- Ex. 1. (...) Posso ter uma recaída ... - **eu** me encarrego de evitar isso.
- Ex. 2. (...) posso compreender por que Glória fez isso - disse **eu**, procurando suavizar a coisa. (...)
- Ex. 3. (...) me referia a você e sim a Glória. - **eu** sei - respondi. - A menina é direita. (...)
- Ex. 4. (...) Glória e eu não dançávamos. **Eu** tinha o braço passado sobre os ombros dela, (...)

Os exemplos 1 e 2 mostram as possibilidades de posição pronominal investigadas. O Ex. 1 mostra um caso em que o pronome ‘eu’ está anteposto ao Processo por ele realizado (‘eu me encarrego’) e o Ex. 2 mostra um caso em que o pronome está posposto ao Processo realizado (‘disse eu’). Já os exemplos 3 e 4 mostram as possibilidades de usos do pronome ‘eu’ (para definir ou enfatizar o sujeito). No Ex. 3, pode-se considerar que o pronome ‘eu’ desempenha função de ênfase do sujeito oracional — uma vez que a desinência verbal (do Processo ‘sei’) permite a identificação indubitável do sujeito. Em casos como o mostrado no Ex. 4, no entanto, o uso do pronome ‘eu’ não apenas enfatiza o sujeito oracional, mas o define — uma vez que a desinência verbal (do Processo ‘tinha’) é comum para a primeira e para a terceira pessoas do sujeito (‘eu tinha’ e ‘ele tinha’).

Feitas as delimitações e classificações apresentadas nos parágrafos anteriores, a planilha na qual os dados estão organizados fica da seguinte forma:

Hit	KWIC	Classificação 1 (Oração)	Interpretação	Classificação 2 ('eu' vs sujeito oracional)	Classificação 3 (Plural vs Singular)	Interpretação	Classificação 4 (posição pronominal)
1	pouco para o lado oposto ao em que eu estava; eu não tinha uma visão completa de	Sim	eu estava do lado x	Sim	Singular	-	Anteposto ao verbo
2	o lado oposto ao em que eu estava; eu não tinha uma visão completa de seu perfil,	Sim	eu não tinha uma visã	Sim	Singular	-	Anteposto ao verbo
3	em agonia? Além disso, ela tinha amigos, sim, eu era o seu melhor amigo. O seu único	Sim	eu era seu amigo	Sim	Singular	-	Anteposto ao verbo
4	sentença não deva ser pronunciada? 2 Que podia eu dizer? ... Toda aquela gente sabia que eu a	Sim	eu podia dizer o que? Sim	Sim	Singular	-	Posposto ao verbo
5	podia eu dizer? ... Toda aquela gente sabia que eu a havia assassinado; a única pessoa que me	Sim	eu a havia assassinad	Sim	Singular	-	Anteposto ao verbo
6	é que fui saber disso. Um dia caminhava eu pela Avenida Melrose, vindo dos estúdios da Par	Sim	eu caminhava pela av	Sim	Singular	-	Posposto ao verbo
7	a quadra, descendo para a Western. - Ah... - fiz eu - pensei que estava acenando pra mim. - A troc	Sim	eu fiz um barulho	Sim	Singular	-	Posposto ao verbo
8	vai sentir prazer em pronunciar a sentença, e eu sinto que essas pessoas atrás de mim vão	Sim	eu sinto alguma coisa	Sim	Singular	-	Anteposto ao verbo
9	vamos tomar aquela manhã em que encontrei Glória. eu não me sentia muito bem; estava ainda um	Sim	eu não me sentia bem	Sim	Singular	-	Anteposto ao verbo
10	filme russo e achei que podia conseguir trabalho. eu costumava perguntar a mim mesmo se haveria cois	Sim	eu costumava pergunt	Sim	Singular	-	Anteposto ao verbo
11	ório central -- desconversou ele, retirando-se. eu não tinha destino certo; estava apenas dirigind	Sim	eu não tinha destino	Sim	Singular	-	Anteposto ao verbo
12	os transeuntes apontarem para mim e dizerem que eu era o maior diretor do mundo... Foi nesse	Sim	eu era o maior diretor	Sim	Singular	-	Anteposto ao verbo
13	com o outro; e quando chegamos à Western eu sabia que o nome dela era Glória Beatty,	Sim	eu sabia o nome dela	Sim	Singular	-	Anteposto ao verbo
14	quarto com outras moças, perto de Beverly, e eu vivia a poucas quadras dali, de sorte que	Sim	eu vivia ali	Sim	Singular	-	Anteposto ao verbo
15	daqui. - Está bem. Ando farta de cinema. Se eu não for melhor atriz do que a maioria	Não	(caso problemático) si-	-	-	-	-
16	melhor atriz do que a maioria daquelas fêmeas, eu como o meu chapéu... Vamos nos sentar no	Sim	eu como meu chapéu	Sim	Singular	-	Anteposto ao verbo
17	sível afastá-las o quanto a gente quisesse ... - eu antes nunca tinha prestado atenção a este lugar	Sim	eu nunca tinha presta	Sim	Singular	-	Anteposto ao verbo
18	tado atenção a este lugar. - Gosto daqui - disse eu, tirando o casaco e estendendo-o sobre a	Sim	eu disse alguma coisa	Sim	Singular	-	Posposto ao verbo
19	ais, mas não consegui registro no Central. - Nem eu. Era preciso a gente estar registrado no Centr	Não	(caso problemático) n-	-	-	-	-
20	o Central Bureau se encarrega de arranjá-los. eu compreendia por que Glória não conseguira regis	Sim	eu compreendia	Sim	Singular	-	Anteposto ao verbo
21	que seja, é melhor que a vida que eu levava lá em casa. - Perguntal -lhe onde vivia	Sim	eu levava a vida	Sim	Singular	-	Anteposto ao verbo
22	o oeste do Texas é um lugar infernal. eu vivia com uma tia e um tio. Ele	Sim	eu vivia com pessoas	Sim	Singular	-	Anteposto ao verbo
23	. Ele era guarda-freios numa estrada de ferro. eu só o via uma ou duas vezes por	Sim	eu o via	Sim	Singular	-	Anteposto ao verbo
24	menos tinha um lar - comentel. - Lar... aquilo? eu lhe dou outro nome. Quando meu tio estava	Sim	eu dou outro nome	Sim	Singular	-	Anteposto ao verbo
25	comigo, quando ele estava trabalhando minha tia e eu sempre vivíamos brigando. Ela tinha medo que eu	Sim	minha tia e eu (nós) vi	Sim	Plural	minha tia e eu (nós)	-
26	eu sempre vivíamos brigando. Ela tinha medo que eu contasse os podres dela ... "Família distinta,	Sim	eu contasse os podres	Sim	Singular	-	Anteposto ao verbo
27	os podres dela ... "Família distinta..." pensei eu. - Assim - continuou Glória -, finalmente fugi	Sim	eu pensei alguma cois	Sim	Singular	-	Posposto ao verbo

Figura 5 — Planilha eletrônica com dados, delimitações e categorias de classificação

Como anteriormente mencionado, a Figura 5, acima, mostra a planilha em que os dados estão organizados. Na coluna D (visível na figura), estão as delimitações dos dados em relação às orações (classificação de sim ou não); na Coluna F, estão as delimitações quanto ao pronome ‘eu’ ser ou não sujeito oracional; na Coluna G estão as classificações relativas ao fato de o pronome construir ideia de plural ou singular e na Coluna I estão as classificações quanto à posição pronominal (anteposta ou posposta ao verbo). As Colunas E e H são reservadas para explicitações das interpretações que levaram às classificações e/ou delimitações (com o simples intuito de facilitar revisões posteriores). Outras colunas (não visíveis na imagem), classificam os usos do pronome ‘eu’ quanto à possibilidade de definição ou ênfase do sujeito e trazem esclarecimentos de outras interpretações de classificações feitas (também com o objetivo de facilitar revisões posteriores nas classificações).

Em relação ao uso de classificações por meio de palavras (diferentemente das classificações por meio de etiquetas numéricas, como exposto na seção 2 deste texto), as planilhas eletrônicas oferecem, por padrão, recursos de auto completar entradas de texto — de forma que ao digitar as primeiras letras de uma classificação anteriormente digitada, a planilha sugere o restante do texto (como mostrado na figura a seguir):

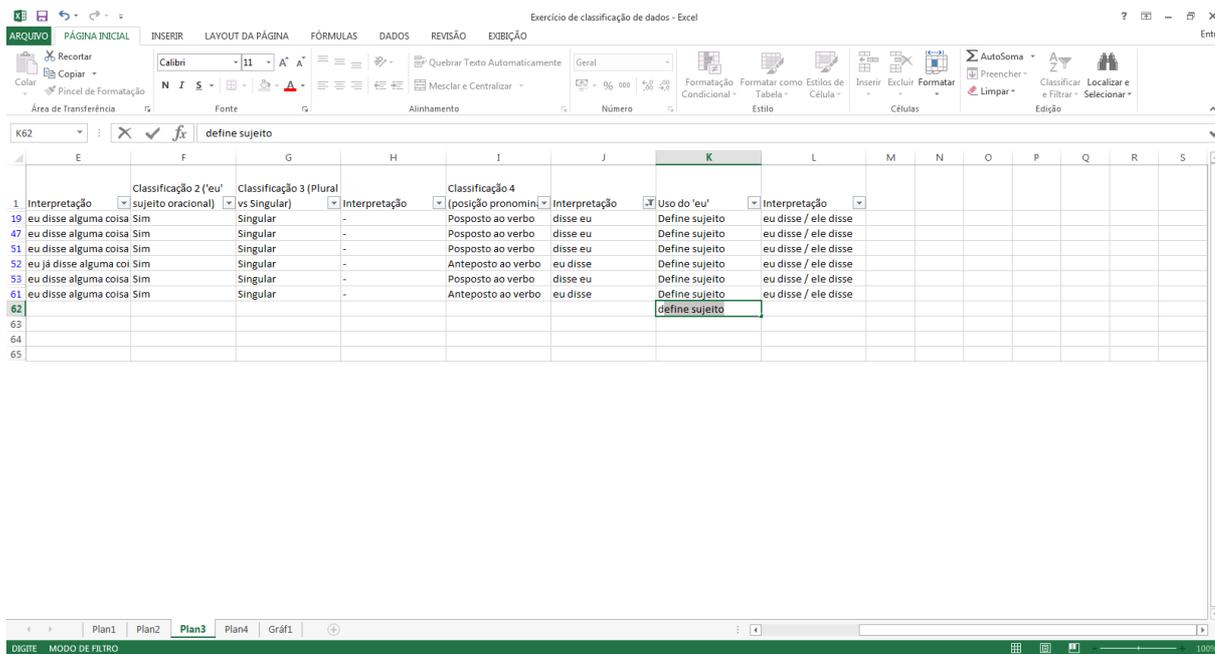


Figura 6 — Uso do recurso auto completar

Como mostra a Figura 6, ao digitar a primeira letra ‘d’ na coluna K (linha 62), a planilha eletrônica sugere o restante do texto ‘efine sujeito’, com base nas classificações feitas anteriormente (bastando um toque na tecla <ENTER> para que o texto fique registrado na célula). O recurso não apenas facilita o processo de classificação como também reduz significativamente a possibilidade de erros de digitação no trabalho de classificação.

Outro recurso — também visível na Figura 6 e também oferecido por padrão em planilhas eletrônicas — que facilita a identificação e a correção de eventuais erros de digitação no trabalho de classificação de dados é o auto filtro. Ativável, no *MS Excel 2013* por meio da guia Página inicial > Classificar e filtrar, o recurso cria listas do tipo drop-list, em que todas as classificações feitas são visualizáveis, o que permite selecionar as células que venham a conter erros de digitação e a correção desses erros de forma rápida, como mostra a figura a seguir:

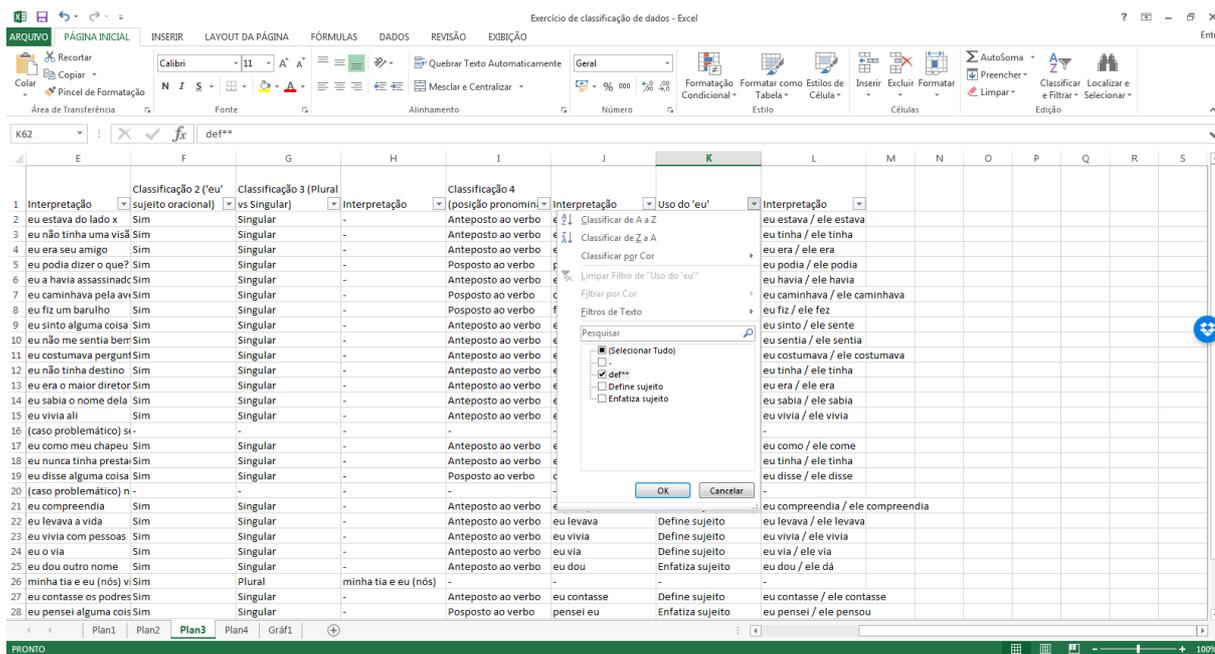


Figura 7 — Identificação de erros de digitação por meio do recurso auto-filtro

Como anteriormente dito, a Figura 7, acima, mostra o processo de identificação de eventuais erros de digitação por meio do recurso auto filtro. Na planilha, foi propositalmente digitada a classificação ‘def**’ (que não corresponde a nenhuma das categorias de análise). A função auto filtro cria a lista *drop-list* (também mostrada na Figura 7). Na lista, é possível selecionar para exibição apenas as classificações que apresentem erros de digitação (e, na sequência, corrigi-los).

Realizados os processos de classificação de dados são utilizadas fórmulas condicionais para contar, somar ou formatar células (e valores) que estejam de acordo com critérios previamente determinados. Nos casos aqui mostrados, é utilizada a função de contagem condicional CONT.SE⁸, de forma que a planilha eletrônica conte quantas vezes ocorre cada segmento textual definido nas classificações.

Para utilizar a função, sugere-se aqui o uso do assistente de funções (acessível por meio do seguinte ícone na barra de ferramentas:

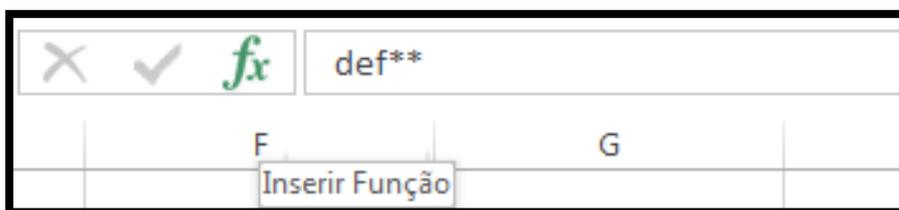


Figura 8 — Ícone para ativar o assistente de funções

Utilizando o assistente de funções, basta localizar a função CONT.SE (mostrada na Figura 9, a seguir) e seguir os procedimentos indicados pelo assistente (mostrados na Figura 10).

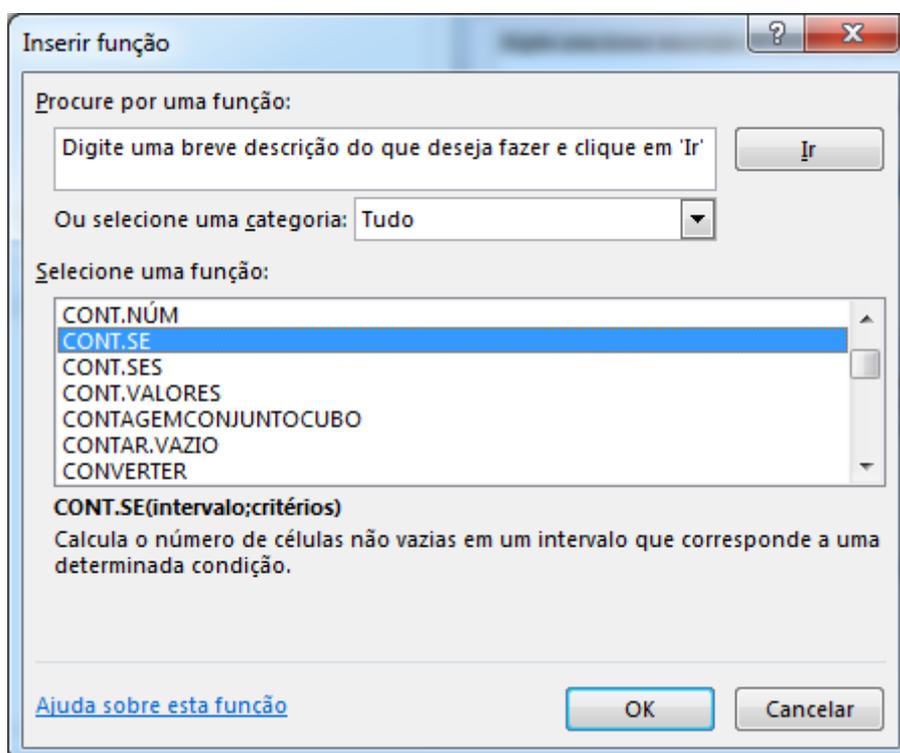


Figura 9 — Assistente de funções (inserindo a função de contagem condicional CONT.SE)

⁸ Dependendo das configurações da planilha, o nome da função pode variar (para Count.If), por exemplo.

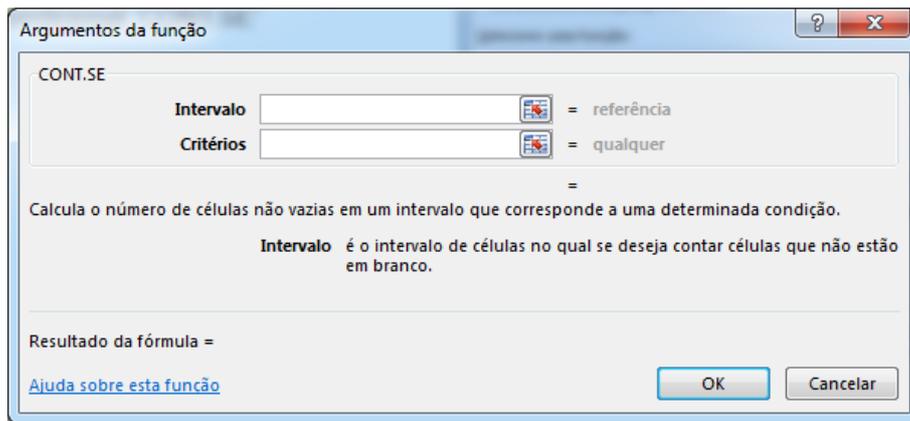


Figura 10 — Argumentos da função de contagem condicional, organizados pelo assistente de funções

Como mencionado anteriormente, acessando-se a função de contagem condicional no Assistente de funções (Figura 9), é aberto o guia para utilização com os argumentos da função em tela (Figura 10). A Figura 11, a seguir, mostra o resultado das classificações aqui realizadas, sendo visível, na barra de ferramentas, a fórmula =CONT.SE(Plan3!D2:D61;"sim"), criada a partir do assistente de funções visto acima. Na fórmula em questão, Plan3!D2:D61 corresponde ao intervalo selecionado para o assistente e "sim" (incluindo as aspas) corresponde ao critério que a fórmula utilizará para contar as células no intervalo.

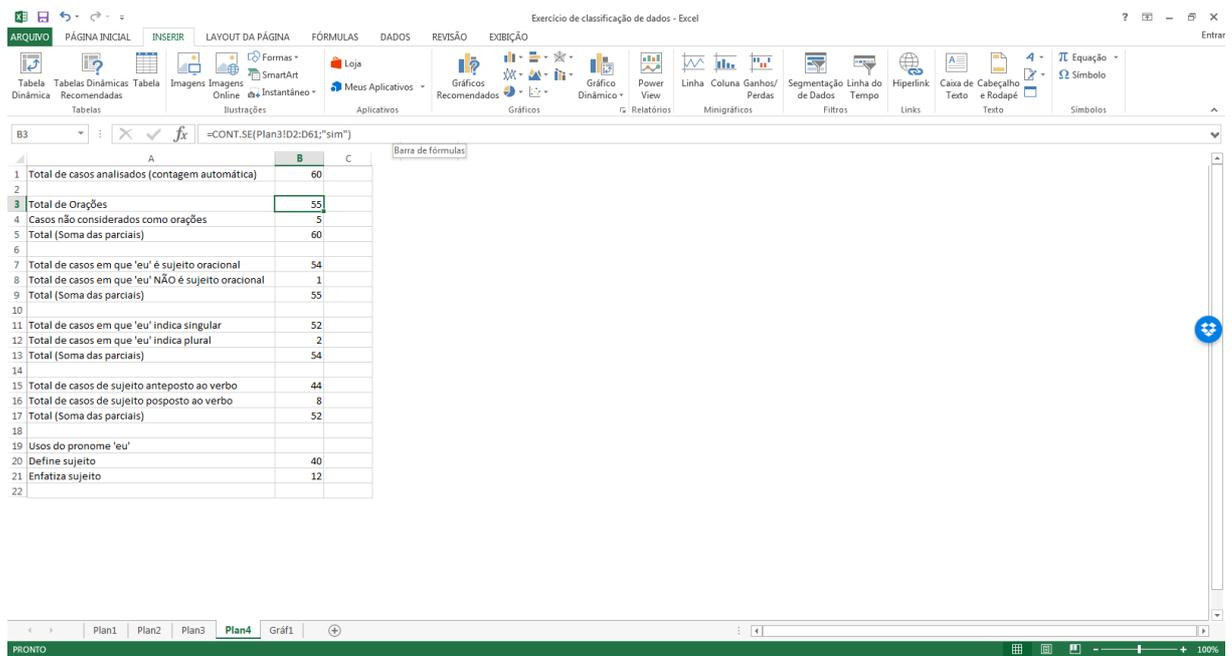


Figura 11 — Resultados da classificação aqui realizada, mostrando, na barra de fórmulas, uma fórmula de contagem condicional

Como anteriormente mencionado, a Figura 11 mostra o das classificações realizadas no estudo desta amostra do corpus. Na figura estão visíveis os resultados das fórmulas de contagem condicional. Anexo a este artigo, está disponibilizada uma tabela que exhibe as fórmulas empregadas na planilha acima, acompanhados de uma breve discussão. Outra vantagem desse método de classificação é a simplicidade de produção de gráficos a partir dos resultados gerados pela planilha (bastando, para tanto, selecionar os dados que comporão o gráfico e, na guia 'Inserir' (visível na imagem acima), escolher um dos assistentes de produção de gráficos e seguir os passos apresentados pelo assistente.

Como o objetivo da discussão apresentada nesta seção é fazer uma indicação metodológica, os resultados do processo de classificação (em termos linguísticos) não serão cotejados com os de Maia (1998). Sugere-se, no entanto, que investigações semelhantes sejam realizadas, retomando as observações da autora e discutindo os usos de pronomes pessoais em português (seja original, seja traduzido).

4. Conclusões e encaminhamentos

A anotação criteriosa do corpus pode acrescentar insumos valiosos para a identificação e interpretação de padrões textuais. Embora haja *softwares* para anotação de corpora, nem sempre eles acomodam as categorias de análise de uma determinada pesquisa, especialmente quando se lida com corpora de pequenas dimensões para análises discursivas. Neste artigo apresentamos duas alternativas de anotação: uma que envolve o desenvolvimento de protocolo ou grade baseada em códigos numéricos para serem inseridos no próprio corpus e outra que usa planilha eletrônica, cuja classificação de dados fica à parte. Cada uma delas apresenta suas vantagens, não são excludentes, podendo até mesmo ser complementares. A inserção de etiquetas no próprio corpus pode ser feita pelo próprio nome da categoria ou através de códigos numéricos atribuídos a cada uma das categorias. Se comparadas entre si, o uso de códigos, ao invés de nomes, reduz a probabilidade de erros de digitação e mantém um visual mais limpo no corpus. A anotação através de planilha eletrônica acresce as facilidades dos recursos deste tipo de ferramenta, especialmente no que se refere ao levantamento estatístico das ocorrências.

Além dos dois métodos aqui apresentados, existem ainda outras opções metodológicas. Citamos - como indicações para pesquisadores(as) que desejam se aprofundar no assunto -, por exemplo, a proposta de Rodrigues (2013) em utilizar a linguagem de marcação extensível XML para anotar e quantificar seus dados e o programa Notepad++ para visualização do trabalho, além das propostas de utilização dos corpora estruturados disponibilizados pelo *Common Crawl*.

Ressaltamos, no entanto, a importância de uma escolha metodológica baseada nos interesses de pesquisa do(a) pesquisador(a) para a obtenção de melhores resultados e sugerimos que sejam investidos tempo e esforço na definição prévia das classificações e métodos (além da realização de estudos piloto para confirmar a viabilidade das escolhas) para reduzir retrabalhos e minimizar atrasos causados por mau planejamento.

Referências

ALVES, D. *Conflito e tradução*: uma análise sobre as realizações linguísticas dos conflitos armados entre grupos litigantes no corpus paralelo Grande Sertão: Veredas — The Devil to Pay in the Backlands. Tese (doutorado) - Universidade Federal de Santa Catarina, 2014 - disponível em: <http://goo.gl/q1B9R7> - último acesso em 23/04/2015.

ALVES, D.; VASCONCELLOS, M. 'A Linguística de Corpus e os Estudos da Tradução: uma análise bibliométrica de pesquisas brasileiras publicadas entre 2006-2010'. *DELTA*. No prelo.

ANTHONY, L. *AntConc3.4.3w*. Tokyo, Japão: Waseda University, 2011 - disponível em: <http://goo.gl/3GVS> - último acesso em 04/01/2012.

FEITOSA, M. P. *Uma proposta de anotação de corpora paralelos com base na lingüística sistêmico-funcional*. 2005. enc. Dissertação (mestrado) - Universidade Federal de Minas Gerais, 2005.

HUNSTON, S. Methods in Corpus Linguistics: beyond the concordance line. In: *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, 2002. p.36-95.

MAIA, B. Word order and the first person singular in Portuguese and English. In: *Meta: journal des traducteurs / Meta: Translators' Journal*, vol. 43, n° 4, 1998. - disponível em: <http://goo.gl/SfzC2F> - último acesso em 23/04/2015.

LEECH, Geoffrey. Adding Linguistic Annotation. In: WYNNE, M. (ed.) *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books, 2005. pp 17-29. Disponível em <http://ahds.ac.uk/linguistic-corpora/>

MARTIN, J. R.; WHITE P. R. R. *The language of evaluation: appraisal in English*. New York: Palgrave Macmillan, 2005.

NOVAIS, Priscila de Oliveira. *A representação de manifestantes e representantes do governo como atores sociais em textos sobre os protestos no Brasil em 2013*. (Mestrado em Linguística) — Programa de Pós Graduação em Linguística — PROLING Universidade Federal da Paraíba, 2015.

RODRIGUES, Roberta Rego. Apresentação do discurso em corpus paralelo literário trilingue. *Tradução & Comunicação: Revista Brasileira de Tradutores*. N. 26. Ano 2013. pp. 93-107 disponível em <http://www.pgsskroton.com.br/seer/index.php/traducom/article/viewFile/1641/1572>.

SINCLAIR, J. M. Preface . In GHADESSY, M et al. (Ed.) *Small corpus studies and ELT*. Amsterdam: John Benjamins, 2001. p. vii-xv.

VAN LEEUWEN, T. The representation of social actors. In: CALDAS-COULTHARD, C. R.; COULTHARD, M. (Eds). *Texts and Practices: readings in Critical Discourse Analysis*. London & New York: Routledge, 1996. p.32-70.

Anexos

A Tabela 1, a seguir, retoma os resultados mostrados na Figura 11 (seção 3 deste artigo), mostrando não apenas os resultados numéricos, mas também as fórmulas empregadas e uma breve discussão sobre suas funcionalidades.

Tabela 1 — Fórmulas utilizadas e resultados do processo de delimitação e classificação de dados apresentado na seção 3

Linha	Resultados	Fórmulas utilizadas	Discussão
Total de Orações	55	=CONT.SE(Plan3!D2:D61;"sim")	A fórmula conta quantas vezes a palavra “sim” ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna D). O resultado (55) indica o número de orações identificadas no intervalo.
Casos não considerados como orações	5	=CONT.SE(Plan3!D2:D61;"não")	A fórmula conta quantas vezes a palavra “não” ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna D). O resultado (5) indica o número de casos em que o ‘eu’ não está em orações (nas linhas de concordância analisadas no intervalo).
Total (Soma das parciais)	60	=SOMA(B3:B4)	Soma as parciais dos resultados exibidos nas células B3 e B4. O total (60) corresponde ao número de linhas selecionadas para esta amostragem.

Linha	Resultados	Fórmulas utilizadas	Discussão
Total de casos em que 'eu' é sujeito oracional	54	=CONT.SE(Plan3!F2:F61;"sim")	A fórmula conta quantas vezes a palavra "sim" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna F). O resultado (54) indica o número de casos em que o pronome 'eu' realiza função de sujeito oracional.
Total de casos em que 'eu' NÃO é sujeito oracional	1	=CONT.SE(Plan3!F2:F61;"não")	A fórmula conta quantas vezes a palavra "não" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna F). O resultado (1) indica o número de casos em que o pronome 'eu' NÃO realiza função de sujeito oracional.
Total (Soma das parciais)	55	=SOMA(B7:B8)	Soma as parciais dos resultados exibidos nas células B7 e B8. O total (55) corresponde ao total de orações identificado nesta amostragem.
Total de casos em que 'eu' indica singular	52	=CONT.SE(Plan3!G2:G61;"singular")	A fórmula conta quantas vezes a palavra "singular" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna G). O resultado (52) indica o número de casos em que o pronome 'eu' constrói ideia de singular.
Total de casos em que 'eu' indica plural	2	=CONT.SE(Plan3!G2:G61;"plural")	A fórmula conta quantas vezes a palavra "plural" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna G). O resultado (2) indica o número de casos em que o pronome 'eu' constrói ideia de plural (como parte de um sujeito composto).
Total (Soma das parciais)	54	=SOMA(B11:B12)	Soma as parciais dos resultados exibidos nas células B11 e B12. O total (54) corresponde ao total de orações em que o pronome 'eu' realiza função de sujeito oracional.
Total de casos de sujeito anteposto ao verbo	44	=CONT.SE(Plan3!I2:I61;"anteposto ao verbo")	A fórmula conta quantas vezes o segmento textual "anteposto ao verbo" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna I). O resultado (44) indica o número de casos em que o pronome 'eu' está anteposto ao verbo.
Total de casos de sujeito posposto ao verbo	8	=CONT.SE(Plan3!I2:I61;"posposto ao verbo")	A fórmula conta quantas vezes o segmento textual "posposto ao verbo" ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna I). O resultado (8) indica o número de casos em que o pronome 'eu' está posposto ao verbo.
Total (Soma das parciais)	52	=SOMA(B15:B16)	Soma as parciais dos resultados exibidos nas células B15 e B16. O total (52) corresponde ao total de orações em que o pronome 'eu' realiza função de sujeito oracional, construindo a ideia de sujeito singular.

Linha	Resultados	Fórmulas utilizadas	Discussão
Define sujeito	40	=CONT.SE(Plan3!K2:K61;"Define sujeito")	A fórmula conta quantas vezes o segmento textual " Define sujeito " ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna K). O resultado (40) indica o número de casos em que o pronome 'eu' é utilizado para definir o sujeito oracional (por esse não ser facilmente recuperável pela desinência verbal).
Enfatiza sujeito	12	=CONT.SE(Plan3!K2:K61;"Enfatiza sujeito")	A fórmula conta quantas vezes o segmento textual " Enfatiza sujeito " ocorre entre as células compreendidas no intervalo entre a linha 2 e a linha 61 (da coluna K). O resultado (40) indica o número de casos em que o pronome 'eu' é utilizado para enfatizar o sujeito oracional (considerando-se que tal sujeito é facilmente recuperável pela desinência verbal).