

A semântica dos verbos de elocução: um estudo a partir do paradigma *FrameNet*

Maristela Maria Kohlrausch¹

¹Programa de Pós-Graduação em Linguística Aplicada – Universidade do Vale do Rio dos Sinos (UNISINOS)

makohlrausch@gmail.com

Resumo. *Este trabalho apresenta os dados preliminares de uma pesquisa de mestrado em Linguística Aplicada que tem como objetivo fazer uma investigação semântico-computacional do léxico do português do Brasil a partir da utilização de corpus eletrônico. A abordagem teórica e metodológica utilizada é a Semântica de Frames (Fillmore 1982, 1985), que está inserida na área da Linguística Cognitiva. Sob a perspectiva da Semântica de Frames, foi criada a base de dados FrameNet, que apresenta as possíveis combinações semânticas e sintáticas (valências) de cada palavra, juntamente com cada um dos seus sentidos, os quais são, nesta pesquisa, os verbos de elocução. Este trabalho apresenta uma proposta de estudo destes verbos a partir deste paradigma e, esta escolha se justifica pelo fato destes verbos serem uma classe que expressa uma série de nuances semânticas, como por exemplo, a subjetividade e a expressão da emoção, permitindo que diferentes perspectivas se voltem para seu estudo. Pretende-se mostrar os resultados de um estudo empírico realizado a partir da extração dos verbos de elocução classificados por Neves (2000) de um corpus constituído por 158 textos. Para a extração utilizou-se uma das ferramentas da Linguística de Corpus, mais especificamente, o concordanciador AntConc 3.2.1w. Como resultado parcial, destaca-se que nem todos os verbos de elocução pertencem, conforme descrição do FrameNet, ao Frame Statement, que seria o equivalente aos verbos de elocução no português. Embora alguns verbos, como “dizer” e “afirmar” tenham apresentado uma descrição semelhante ao do FrameNet, ocorreram casos, como “argumentar” e “considerar”, que não apresentaram uma relação direta com o Frame Statement. O que se pode perceber foi que, ao comparar duas línguas, questões como os diferentes níveis de granularidade e paralelismos entre as mesmas são totalmente relevantes. Este estudo contribui para o desenvolvimento de sistemas computacionais que necessitam processar a linguagem humana, em especial à pesquisadores do Processamento da Linguagem Natural (PLN).*

Abstract. *This paper presents a preliminary study from a Master’s research in Applied Linguistics which aims to make a computer-semantic Portuguese lexicon investigation from the use of electronic corpus. The theoretical and methodological approach used is Semantics Frames (Fillmore 1982, 1985), which is inserted in the field of Cognitive Linguistics. In the perspective of Semantics Frames was established the FrameNet database, which presents the possible semantic and syntactic (valences) combination for each word, with*

each of its senses, which are in this search, the verbs of elocution. This paper presents a proposal for a study of these verbs from this paradigm, and this choice is justified because these verbs are a class that expresses a series of semantic nuances, such as the subjectivity and emotion expression, allowing different perspectives related to its study. It is intended to show the results of an empirical study carried out from the extraction of the verbs of elocution classified by Neves (2000), of a corpus composed by 158 texts. For the extraction one of the Corpus Linguistics tools was used, namely a concordancer called AntConc 3.2.1w. As a preliminary result, we point out that not all verbs of elocution belong, according to description of FrameNet, to the Frame Statement, which would be equivalent to the verbs of elocution in Portuguese. While some verbs such as “dizer” and “afirmar” have presented a similar description to FrameNet, there were cases as “argumentar” and “considerar”, which did not show a direct relationship with the Frame Statement. We have noticed that, when comparing two languages, issues such as different levels of granularity and parallelism between them are totally relevant. This study contributes to the development of computer systems that need to process human language, especially the researchers of the Natural Language Processing (NLP).

Palavras-chave: semântica de *frames*; *FrameNet*; verbos de elocução

1. Introdução

O presente trabalho apresenta os dados preliminares de uma pesquisa de mestrado em Linguística Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS, em São Leopoldo/RS, cujo principal objetivo é fazer uma investigação semântico-computacional do léxico do português do Brasil a partir da utilização de *corpus* eletrônico.

O foco da pesquisa foram os **verbos de elocução**, que, em linhas gerais, são verbos que introduzem o discurso e estabelecem um elo entre enunciados de diferentes enunciações em virtude da sua variedade e riqueza em matizes semânticos. O objetivo foi verificar em que medida a teoria de *frames* (Fillmore 1982, 1985) – que parte do princípio de que, para entender os conceitos que caracterizam pequenas cenas abstratas ou situações, é necessário considerar a estrutura completa em que estas situações se encontram – se presta à descrição semântica desta classe verbal, ou seja, em que medida a semântica de *frames* possui adequação descritiva e explanatória para dar conta dos dados em Língua Portuguesa. Para isso, foi realizado um estudo semântico destes verbos a partir do paradigma *FrameNet*¹ – um projeto computacional lexicográfico baseado em *corpus* – tendo em vista os diferentes padrões de lexicalização entre inglês e português.

Pesquisas neste âmbito são relevantes, pois há uma escassez de investigações voltadas para a construção de recursos computacionais com base em estudos semânticos no Brasil. Vários estudos vêm sendo realizados em diferentes países sob esta perspectiva. O projeto FN, por exemplo, possui versões em diferentes línguas, como

¹ No decorrer do trabalho utilizarei a sigla FN quando me referir ao *FrameNet*.

alemão, espanhol e japonês. Além disso, a escolha pelos verbos de elocução foi feita, além dos motivos citados acima, pela recorrência desta classe no *corpus* de estudo.

O trabalho está dividido da seguinte maneira. Esta seção, a introdução, onde apresento a pesquisa, juntamente com os objetivos e justificativas. Na seção 2, apresento o foco do trabalho e o porquê da escolha pelos verbos de elocução. A seguir, na seção 3, questões metodológicas são abordadas. Na seção 4, apresento a análise dos dados. Finalmente, na seção 5, as conclusões preliminares deste estudo.

2. A semântica dos verbos de elocução

O foco deste trabalho foi na semântica dos verbos de elocução do português brasileiro. A escolha por esta verbal se deve ao fato desses verbos serem recorrentes no *corpus*, além de expressarem uma série de nuances semânticas, como, por exemplo, a subjetividade e a expressão da emoção, o que explica o fato de serem abordados por teorias que seguem princípios distintos. Além disso, optou-se por este enfoque em virtude destes verbos carregarem informações de natureza discursiva.

Em linhas gerais, estes verbos são usados pelo discurso direto e indireto. Tradicionalmente conhecidos, no inglês, como *Reported Speech*, esta forma de reproduzir a fala engloba o discurso direto e indireto, pois depende da maneira como o locutor principal reproduz a voz de outros falantes. Esta opção por parte do locutor nunca é aleatória, assim como também não é aleatória a escolha do verbo associado à reprodução da fala. No discurso direto, por exemplo, o narrador reproduz a fala do personagem, por meio de verbos de elocução, e isso faz com que este tipo de discurso valorize o falante e revele maior isenção do locutor principal. Já no discurso indireto, o narrador conta o que o personagem fala, ou seja, ele é testemunha auditiva e reproduz para o leitor o que ouviu do personagem. Embora o discurso indireto tenha maior interferência do locutor principal, a presença dos verbos de elocução é imprescindível.

Para este estudo, foram selecionados os verbos de elocução classificados por Neves (2000), que sob uma perspectiva funcionalista, categoriza-os como sendo verbos introdutores de discurso (tanto o direto como o indireto) e classifica-os em **(a)** verbos de *dizer*, **(b)** verbos que *qualificam o que é dito* e **(c)** verbos que *circunstanciam ou instrumentalizam o que é dito*. Os verbos de *dizer*, segundo ela, são os verbos de elocução propriamente ditos, ou seja, verbos de ação cujo complemento direto é o conteúdo do que se diz. Incluem-se neste grupo os verbos *falar, dizer, gritar, exclamar, repetir*. No grupo dos verbos *que qualificam o modo que caracteriza o dizer* fazem parte os verbos *comentar, queixar-se, explicar, protestar* e no grupo *que circunstanciam ou instrumentalizam o que foi dito*, estão os verbos que introduzem o discurso, mas não necessariamente indicam atos de fala. Fazem parte os verbos *acalmar, consolar, rir, chorar, suspirar*.

Embora os verbos de elocução possam ser vistos sob distintas perspectivas, este estudo inspirou-se nos verbos classificados por Neves (2000) em virtude da abrangência abordada pela autora.

3. Procedimentos metodológicos

Para a seleção dos dados, optou-se por utilizar o *Corpus* PLN-BR FULL², especificamente a seção Cotidiano, que possui 158 textos, totalizando 42378 *tokens* e 8724 *types*. Os textos são dos cadernos do jornal Folha de São Paulo e variam do ano de 1994 a 2005. Este *corpus* foi o ponto de partida, pois serviu como base para a extração dos dados e identificação e análise dos *frames*.

O intuito, com esta busca, foi verificar a quais *frames* os elementos evocadores de *frames*, ou seja, os verbos de elocução pertenciam. Isso se deve ao fato dos verbos serem uma das principais unidades lexicais evocadores de *frames*. No FN, as palavras estão relacionadas a *frames* semânticos dos quais elas participam; e os *frames*, por sua vez, estão relacionados a uma lista de palavras e a outros *frames* com que tenham alguma afinidade semântica. Sob esta perspectiva, os *frames* contêm a relação de participantes ou argumentos que podem compor a situação ou evento por eles descritos. Os participantes (argumentos) são chamados na base de dados de *elementos frames*³. Além disso, cada *frame* apresenta uma lista de *unidades lexicais*⁴ que se relacionam ao *frame* em questão. Essas ULs são chamadas de *elementos evocadores de frame*. Uma UL é a combinação de um conceito semântico com a informação morfossintática, e que pode ser descrita com os EFs próprios do *frame* ao qual pertence. Partiu-se dos verbos classificados por Neves (2000) como sendo verbos de elocução.

Após a seleção dos verbos de elocução, foi feita a busca destes verbos no *corpus*. Para a extração e seleção dos dados do corpus Cotidiano, foi utilizado um concordanciador chamado *AntConc* 3.2.1w⁵, uma ferramenta que facilita a coleta e análise dos dados. Ele é um *software* gratuito desenvolvido por Laurence Anthony da Universidade da Waseda, no Japão. O programa está disponível no *site* do pesquisador gratuitamente e permite fazer vários tipos de pesquisa em um determinado *corpus*. Não é necessário fazer nenhuma instalação, pois o programa consiste em apenas um arquivo (de 3,67MB) que é executado com um duplo clique de *mouse*. O programa é fácil de usar, porém ele não identifica as conjugações dos verbos, e, como o foco foi nos verbos de elocução, foi preciso pesquisar palavra por palavra e identificar os verbos com suas possíveis conjugações.

Selecionados os verbos de elocução, ou seja, as ULs evocadoras de *frame*, foi feita a consulta na base de dados do FN para verificar quais serão os *frames* evocados pelos verbos estudados. Para isso, precisou-se procurar um equivalente em inglês para estes verbos em virtude da base de dados do FN ser nesta língua.

A próxima etapa da análise consistiu em descrever os verbos de elocução com base nas categorias conceituais oferecidas pela base de dados do FN. Esta fase incluiu não apenas a descrição dos *frames*, mas também os EFs e as relações entre os *frames*,

² Este *corpus* faz parte do projeto PLN-BR que tem como objetivo a construção de recursos e ferramentas para a recuperação de informação em bases textuais em português do Brasil.

³ No decorrer do trabalho utilizarei a sigla EFs quando me referir aos *elementos frame*.

⁴ No decorrer do trabalho utilizarei a sigla ULs quando me referir às *unidades lexicais*.

⁵ Disponível em <<http://www.antlab.sci.waseda.ac.jp/software.html>>

pois foi feita a comparação do português com o inglês, abordando os problemas de lexicalização entre as línguas. A seguir, os dados preliminares são apresentados.

4. O estudo comparativo

Nesta seção apresento alguns dados encontrados. Destacam-se questões relacionadas aos diferentes padrões de lexicalização, bem como a correspondência e paralelismo entre o português e inglês.

4.1 Diferenças nos padrões de lexicalização / correspondência

Após a extração dos verbos de elocução, com a ajuda do concordanciador *Antconc 3.2.1w*, foi feita a consulta na base de dados do FN para verificar quais os *frames* evocados pelos verbos em questão. Nesta etapa, foram consultados 103 *elementos evocadores de frames*. Vale esclarecer que somente alguns foram analisados para este artigo. A consulta foi feita, em um primeiro momento, partindo do verbo correspondente em inglês. Em virtude da base de dados do FN ser em inglês, foi necessário encontrar um equivalente dos verbos de elocução. Para isso foram utilizados dicionários bilíngües *Oxford* e *Websters* português-inglês. Neste momento, questões como a polissemia, informações valenciais e os diferentes padrões de lexicalização foram enfrentados.

A título de exemplo, os verbos de elocução *dizer* e *considerar* não tiveram problemas no que diz respeito à correspondência/paralelismo. Foi encontrado como equivalentes os verbos *say* e *consider*, respectivamente. Já os verbos de elocução *afirmar* e *informar* tiveram mais de uma unidade lexical que expressasse o mesmo sentido para o inglês. Foram encontrados os verbos *affirm*, que segundo o dicionário monolíngüe *Oxford Advanced Learner's* a descrição deste verbo se dá pela “declaração que algo é verdadeiro”; e *state*, que segundo o mesmo dicionário, é “escrever ou dizer formalmente algo, especialmente de forma cuidadosa e clara”, como equivalentes ao verbo *afirmar*. Para o verbo *informar*, foram encontrados os verbos *report*, que é definido como “dar uma descrição sobre algo ou informar sobre algo para alguém” e *inform*, que segundo o dicionário é “contar para alguém sobre algo, especialmente de uma maneira oficial”. Embora os significados destas ULs e de todas as outras que são classificadas como verbos de elocução não sejam diferentes no que diz respeito ao contexto, vários verbos tiveram mais de uma unidade lexical no inglês que expressasse o mesmo significado no português.

4.2 Os verbos de elocução e o *frame Statement*

Após a etapa de encontrar um equivalente para os verbos no português, iniciou-se a identificação dos verbos de elocução no FN. O objetivo foi verificar a quais *frames* estas unidades lexicais pertencem. Ao pesquisar no FN as unidades lexicais *negar*, *alegar*, *declarar*, *reafirmar*, *dizer*, *exclamar*, verificou-se que elas pertencem ao *frame Statement*, que segundo o FN é um *frame* que contém verbos que comunicam o ato de um *Speaker*⁶ direcionando uma *Message*⁷ para algum *Addressee*⁸ usando linguagem.

⁶ remetente/falante

⁷ mensagem

Possui como **elementos frames centrais** *medium, message, speaker, topic*⁹ e **elementos frames não-centrais** *addressee, manner, place, time*¹⁰. Para cada elemento *frame*, está disponível um exemplo com sentenças anotadas para cada evento.

O estudo mostrou que as unidades lexicais evocadas pelo *frame Statement* em português têm um comportamento similar com as ULs em inglês. As informações semânticas podem ser usadas para a descrição das sentenças no português. Encontramos os **elementos frames** *speaker, medium e message* em nosso *corpus*. As informações sintáticas também são similares, como podemos ver abaixo:

```
[<Message> <QUO> "Há tendência de queda"], disse
[<Speaker> <DNI>].

[<Message> <QUO> "O que nós fizemos foi um
ajuste fino], disse [<Speaker> <NP.Ext> o
pesquisador] [<Addressee> <PP> à Folha], [<Medium>
<PP> por telefone].

[<Speaker> <NP.Ext> Ele] disse [<Message> <Sinf> que
a ministra Marina Silva criou um grupo de
trabalho para analisar as propostas].
```

Figura 1: Sentenças anotadas em português

Como se pode ver, a sintaxe e a semântica das sentenças em português com o verbo *dizer* geralmente correspondem ao inglês. Os padrões de lexicalização geralmente não mudam. Esta evidência significa que o *frame Statement* pode ser usado para anotar sentenças do português com este verbo sem dificuldades significativas. Para este artigo, o detalhamento da anotação das sentenças, bem como a ferramenta utilizada não serão abordados.

4.3 Os verbos de elocução e os frames relacionados ao *Statement*

Embora alguns dos verbos de elocução tenham relação somente com o *frame Statement*, muitas outras unidades lexicais pertencem a outros *frames* que possuem alguma relação, mesmo que indireta, com o *frame* em questão.

É o caso do verbo *argumentar*, do inglês *argue*. Ao pesquisarmos no FN verificamos que esta UL evoca o *frame Quarreling*¹¹, que é herdado do *Discussion*¹²,

⁸ destinatário.

⁹ meio, mensagem, falante e assunto.

¹⁰ destinatário, maneira, lugar, tempo.

¹¹ Este *frame* descreve uma situação onde um grupo de argumentadores incompatíveis expressam opiniões ou crenças sobre um assunto.

que usa o *Communication*¹³. O *frame Statement* é herdado do *frame Communication*. *Quarreling*, que também é uma forma de comunicação, apresenta uma relação indireta com o *Statement*. Já com o verbo **contar**, do inglês **tell**, não ocorreu o mesmo. Na base de dados do FN, esta UL pertence ao *frame Telling*¹⁴, que é herdado do *Frame Statement*. Outro exemplo é o verbo **citar**, do inglês **cite**. Esta UL evoca o *frame Adducing*¹⁵, que possui relação com o *frame Statement* no que diz respeito ao uso, ou seja, possui uma relação mais aberta entre este *frame*, sem serem necessariamente herdados um do outro. O que se pode perceber é que, embora as ULs estejam relacionadas a situações de comunicação, elas não são descritas em um único *frame*.

4.4 Os verbos de elocução sem nenhuma relação com o *frame Statement*

Embora muitas ULs estudadas tenham apresentado um comportamento similar ao do inglês, fazendo parte do *frame Statement* ou dos *frames* herdados ou inerentes do *Statement*, ocorreram casos em que os verbos de elocução não tiveram nenhuma relação com os *frames* de comunicação.

Foi o caso do verbo **consider**, do português **considerar**, classificado por Neves (2000) como um verbo de elocução. Este verbo não consta na lista das ULs evocadoras do *frame Statement*. Na base de dados do FN, o EF **consider** evoca o *frame Categorization*¹⁶ e o *frame Cogitation*¹⁷, não sendo nenhum dos *frames* herdados ou inerentes do *Statement*. O *frame Categorization* possui uma relação de uso, totalmente indireta com o *frame Communicate_categorization*, mas este, por sua vez, só possui relação de uso com o próprio *frame Categorization*. O *frame Cogitation*, outro *frame* evocado pela unidade lexical **consider**, não possui nenhuma relação com *frames* de comunicação.

¹² Este *frame* descreve uma situação onde duas ou mais pessoas falam uma com a outra. Nenhuma pessoa é interpretada como sendo o falante ou o destinatário. Ao contrário, ambos falam ou ouvem (o processo da situação é entendido por ser simétrico ou recíproco).

¹³ Este *frame* descreve uma situação onde um comunicador transmite uma mensagem para um destinatário. O assunto e o meio da comunicação também podem ser expressos. Este *frame* não inclui especificações do método da comunicação (fala, escrita, gesto, etc.).

¹⁴ Este *frame* descreve uma situação onde um falante dirige-se a um destinatário com uma mensagem, que pode ser indiretamente referida a um assunto.

¹⁵ Este *frame* descreve uma situação onde um falante menciona uma entidade específica desempenhando um papel, fornecendo assim informações que suportem o ponto de vista do falante.

¹⁶ Este *frame* descreve uma situação onde um *cognizer* constrói um item como pertencente a uma certa categoria. Neste processo, o *cognizer* pode perceber passivamente o item e notar que ele se enquadra aos critérios de uma categoria, ou, alternativamente, analisar o item como um critério que define uma categoria.

¹⁷ Este *frame* descreve uma situação onde uma pessoa, o *cognizer*, pensa sobre um tema durante um período de tempo. O que está pensando pode ser um curso de ação que uma pessoa pode ter, ou algo mais geral.

5. Considerações Finais

Como se pode perceber, esta simples análise evidencia um dado interessante: as unidades lexicais analisadas - *negar, alegar, declarar, reafirmar, dizer, exclamar* – são descritas em um único frame de comunicação, podendo ser classificadas como verbos de elocução neutros, com total relação com o *frame Statement*. Outras unidades lexicais analisadas – *argumentar e contar* -, ainda que estejam relacionadas a situações de comunicação, não são descritas em um único *frame*. Já a unidade lexical – *considerar* – por exemplo, não apresentou nenhuma relação com os *frames* de comunicação descritos no FN.

O que se pode concluir, diante dos dados apresentados, é que o FN não apresenta a classe “verbos de elocução” da mesma forma como muitos dos estudos clássicos que tratam do tema. Contudo, isso não implica dizer que o FN não dá conta da semântica desses verbos. Ao propor uma diversidade de *frames* a partir de distintas situações, o FN propõe uma perspectiva de análise distinta para tais verbos. Resta avaliar, a partir dos dados já obtidos, se tal proposta é descritiva e explanatoriamente adequada para os verbos de elocução.

6. Bibliografia

ATKINS, S., RUNDELL, M., SATO, H. The Contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography*. Vol. 16, 333-357, 2003.

BASE DE DADOS LEXICAIS PARA A LÍNGUA INGLESA – FrameNet. Universidade da Califórnia em Berkeley: The International Computer Science Institut (ICSI), disponível em <http://www.icsi.berkeley.edu/~FrameNet> > .

BOAS, H. C. Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography*. Vol. 18, 445-478, 2005.

EAGLES. EAGLES LE3-4244: *Preliminary Recommendations on Semantic Encoding*. Final Report, 1999.

FERRAMENTA AntConc 3.2.1w: concordanciador, disponível em <<http://www.antlab.sci.waseda.ac.jp/software.html>>.

FILLMORE, Charles J. Frame semantics. In: *Linguistics in the Morning Calm*, Seul: Hanshin Publishing Co., p.111-137, 1982.

_____. Frames and the semantics of understanding. *Quaderni di Semantica*, v.6, n.2, p. 222-254, 1985.

_____.JOHNSON, Christopher R.; PETRUCK, Miriam R.L. Background to FrameNet. *International Journal of Lexicography*. Vol.16, 235-250, 2003.

FONTENELLE, Thierry (2003). FrameNet and Frame Semantics. *International Journal of Lexicography*. Vol. 16. Oxford University Press.

- LÖNNEKER - RODMAN, Birte L. Multiliguality and FrameNet. *International Computer Science Institut*. Berkeley, California, 2007.
- NEVES, M. H. M. *Gramática de usos do português*. São Paulo:UNESP, 2000.
- OHARA, Kyoko Hirose; FUJII, Seiko. Frame Semantics and FrameNet (In Japanese). *The Rising Generation*. Vol.149. No.6: 373-376,387, 2003.
- OXFORD Advanced Learner's Dictionary. Disponível em: <
<http://www.oup.com/elt/catalogue/teachersites/oald7/?cc=global>>
- PADÓ, Sebastian. Cross-lingual Annotation Projection Models for Role-Semantic Information. Saarbrücken: Universität des Saarlandes, 2007.
- SUBIRATS, Carlos; PETRUCK, Miriam. Surprise: Spanish FrameNet. *International Congress of Linguists*. Workshop on Frame Semantics, Prague (Czech Republic), July 2003.
- WEBSTERS. *Dicionário Inglês-Português*. Nova edição. Rio de Janeiro; São Paulo: Record, 2001.