

A polissemia dos verbos psicológicos: um estudo comparativo entre o Espanhol e o Português com base no paradigma *FrameNet*

Silvia Matturro Panzardi Foschiera ¹

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

silvia.matturro@hotmail.com

Resumo. O tema do projeto a ser desenvolvido é A polissemia dos verbos psicológicos: um estudo comparativo entre o Espanhol e o Português com base no paradigma *FrameNet*. Encontra-se circunscrito na área da Lingüística Cognitiva e inserido na linha de pesquisa Texto, Léxico e Tecnologia que constitui um dos campos de investigação do programa de Pós-Graduação em Lingüística Aplicada da Universidade do Vale do Rio dos Sinos. Com essa investigação se busca desenvolver um estudo comparativo entre as línguas espanhola e portuguesa de vertente brasileira, abordando a polissemia dos verbos psicológicos e utilizando como base a Semântica de *Frames*. Tendo como objetivo estudar os padrões de lexicalização entre espanhol e português dos verbos de processamento mental, será possível investigar e identificar quais os itens lexicais que possuem padrões diversos nestas línguas, buscando propor uma nova classificação que atenda as características de ambas as línguas.

Palavras-chave: Lingüística Cognitiva; Lingüística Computacional; Semântica de *Frames*; verbos psicológicos; língua espanhola; língua portuguesa.

Resumen. El tema del proyecto que será desarrollado es *La polisemia de los verbos psicológicos: un estudio comparativo entre el Español y el portugués teniendo como base el paradigma FrameNet*. Está circunscrito en el área de la Lingüística Cognitiva e insertado en la línea de investigación Texto, Léxico y Tecnología, que constituye uno de los campos de investigación del programa de Postgrado en Lingüística Aplicada de la *Universidade do Vale do Rio dos Sinos*. Con esta investigación se busca desarrollar un estudio comparativo entre las lenguas española y portuguesa de vertiente brasileña, abordando la polisemia de los verbos psicológicos, y utilizando como base da Semântica de *Frames*. Teniendo por objetivo estudiar los patrones de lexicalización entre español y portugués de los verbos de proceso mental, será posible investigar e identificar cuales ítems lexicais poseen patrones diversos en estos idiomas, buscando proponer una nueva clasificación que atienda las características de ambas lenguas.

Palabras-clave: Lingüística Cognitiva; Lingüística Computacional; Semântica de *Frames*; verbos psicológicos; lengua española; lengua portuguesa.

1. Projeto

Na atualidade, o processamento da linguagem natural é objeto de interesse da Lingüística, da Computação e da Lingüística Computacional, configurando-se como uma área de pesquisa multi e interdisciplinar. O estudo da polissemia verbal vem sendo fomentado através do advento dos tradutores automáticos, das redes semânticas e dos serviços de busca através da Internet.

Em língua espanhola, na área da Lingüística Computacional, a pesquisa sobre o léxico e suas relações vem recebendo destaque. Na língua portuguesa de variante brasileira as iniciativas ainda são de pequeno número. No Brasil, na área da Lingüística, a quantidade de investigações sobre polissemia ainda é bastante pequena. Este número se torna menor ainda se os estudos forem realizados à luz de teorias lexicais de natureza cognitiva. Já, nos países de língua espanhola, em especial na Espanha, o número de investigações é maior. Na atualidade encontram-se em desenvolvimento várias pesquisas sobre a polissemia, a maioria delas dentro da área da Lingüística Computacional, e motivadas, em grande parte, pelo projeto *Spanish FrameNet*, sediado na Espanha.

O objeto de investigação deste projeto são os verbos psicológicos nos idiomas português e espanhol, que, em função de sua característica, oferecem uma dificuldade bastante grande para serem tratados por tradutores automáticos. Os verbos psicológicos são aqueles verbos que designam ações, estados ou processos anímicos, mentais ou cognitivos e, nos quais, o participante envolvido é um experienciador e não um agente. Segundo C. MELLIS (1999) são verbos que denotam um processo de afetação psicológica nos quais figuram o ser animado que sofre o processo, conhecido como “experimentante”, e a entidade externa que suscita o processo no experimentante, chamada de “estímulo”.

Os verbos psicológicos apresentam um comportamento singular, tanto sintático quanto semântico, e características especiais na estrutura temática e na estrutura argumental, se comparados com outros tipos de verbos. Neles, o experienciador pode aparecer tanto na posição de sujeito como na posição de objeto e, no caso dos verbos transitivos, a anáfora só funciona conforme a regra, isto é comandada pelo antecedente, para os casos em que o verbo tem o experienciador na posição de objeto.

Essa condição fez deles objeto de estudo em diversas línguas, como é o caso dos trabalhos de Levin (1993) na língua inglesa, de Belletti y Rizzi (1987) e Saltarelli (1992) no italiano, de Ruwet (1994, 1995a e 1995b) e Bouchard (1992) no francês, de Caçado (1999) no português e de Whitley (1995 e 1998) e García-Miguel (1995) no espanhol. Estes verbos também foram analisados sob diversas perspectivas teóricas, tais como a gramática relacional (Legendre 1989), a gramática gerativa (Herschensohn 1992, Zubizarreta 1992, Johnson 1992, Pesetsky 1987), a teoria localista (van Voorst 1992 e 1995), o modelo lexemático funcional (Faber e Mairal Usón 1999), e a teoria do léxico-gramática (Mathieu 2000),

No desenvolvimento deste projeto, será adotada uma perspectiva semântico-conceitual e relacional na classificação. A reflexão estará sustentada nas seguintes linhas: (1) o léxico e a semântica lexical; (2) a polissemia verbal de verbos de psicológicos em estudos lingüísticos e computacionais; (3) as bases de dados fundamentadas na semântica de *Frames*.

2. Fundamentos teóricos

Este projeto encontra-se circunscrito teoricamente dentro da área da Linguística Cognitiva, área que vem se consolidando nos últimos quinze anos em função do crescente aumento no número de investigações. Segundo Silva (2006), a Linguística Cognitiva se constituiu como paradigma científico após a realização, na Alemanha, do primeiro *International Cognitive Linguistics Conference*, em 1989. Os primeiros estudos surgiram nos anos 80, a partir de lingüistas como G. Lakoff, R. Langacker e L. Talmy. É uma área de tradição funcionalista, originada dos estudos do significado, que se coloca em oposição à Semântica Gerativa de tradição formalista.

De acordo com Silva (2004), para a Linguística Cognitiva a linguagem é parte integrante da cognição e não um módulo separado. A linguagem é o objeto da relação epistemológica, cabendo à área investigar como esse conhecimento é adquirido. Para isso, fundamenta-se em processos cognitivos, sócio-interacionais e culturais, estudando a linguagem no uso e no contexto da conceitualização, da categorização, do processamento mental, da interação e da experiência individual, social e cultural. Em contrapartida, a Linguística Gerativa assume esses mesmos fatores como secundários ou auxiliares.

A Linguística Cognitiva encontra-se caracterizada por três princípios fundamentais: o da primazia da semântica na análise lingüística e de natureza enciclopédica e perspectivista do significado lingüístico; o da significação como fenômeno lingüístico primário e o da linguagem como meio de interpretar e construir o conhecimento do mundo.

Fundamentada na Linguística Cognitiva, a Semântica de *Frames* é um modelo de representação do conhecimento, de estruturação do léxico, desenvolvido por C. Fillmore (1982, 1985). Esse modelo está baseado na noção de *frame* semântico que é uma estrutura conceitual evocada ou codificada em uma construção na qual a palavra é a construção léxica mais básica.

Segundo Subirats (2006), a Semântica de *Frames* apresenta como princípio que o significado das unidades lexicais se constrói tendo como suporte um conjunto de conhecimentos que deve ser analisado em função dos marcos semânticos que evocam ditas unidades. Entende-se por marco semântico uma representação esquemática de uma situação na qual estão imbricados vários participantes, objetos de cenário e outros papéis conceituais, que, no seu conjunto, constituem os argumentos semânticos de tal *frame*. No marco teórico da Semântica de *Frames*, as palavras não se relacionam umas com as outras, mas o fazem através dos *frames* que servem de suporte comum e que atuam de acordo com a base conceitual para determinar o seu significado.

A Semântica de *Frames* é a teoria através da qual se fundamenta o projeto *FrameNet*. Este projeto consiste na geração de um léxico da língua inglesa que contém as possibilidades combinatórias sintáticas e semânticas de cada unidade léxica. Para isso, as propriedades sintáticas e semânticas das palavras são relacionadas com os *frames* semânticos nos quais tais unidades léxicas estão inscritas.

O *FrameNet* integra uma investigação mais ampla, formada por vários grupos de pesquisa: o *FrameNet2* (Estados Unidos), que deu início aos trabalhos, o *Spanish FrameNet3* (Espanha), segundo a ser estruturado, o *German FrameNet* e o *SALSA II4* (Alemanha), e o *Japanese FrameNet5* (Japão). O objetivo destes projetos é desenvolver

recursos léxicos com anotação semântica e sintática de ampla cobertura léxica e utilizar os recursos léxicos gerados por cada um como um corpus de teste de aplicações de etiquetagem automática de tipos semânticos. Os quatro projetos têm em comum o mesmo modelo semântico e o mesmo formato, o que permite comparar suas experiências lingüísticas e suas metodologias, compartilhar o software de anotação semântica e de etiquetagem semântica automática.

Na investigação pretendida, o léxico computacional permitirá a realização de um mapeamento quanto à informação relativa à estrutura semântica e sintática dos verbos psicológicos nos idiomas espanhol e português, utilizando respectivamente os léxicos semântico-computacionais *Spanish FrameNet* e *Frame Corp*.

O projeto de pesquisa *Spanish FrameNet*, financiado pelo Ministério de Educação da Espanha está sendo desenvolvido conjuntamente pela Universidade Autônoma de Barcelona e o *International Computer Science Institute* de Berkeley. O projeto *Spanish FrameNet* é uma resposta para o desenvolvimento de sistemas de processamento semântico automático da informação textual em espanhol com os fins mais diversos, já seja para sistemas de pergunta-resposta, extração automática de informação, ajuda à tradução, entre outros.

O objetivo dele é aplicar a teoria da semântica de *Frames* para investigar os verbos, os nomes e adjetivos predicativos, os advérbios, as preposições, as conjunções e os nomes de entidades. Na teoria utilizada no estudo da semântica de *Frames*, o significado das unidades léxicas se constrói a partir de um fundo de conhecimentos, que devem ser analisados em função dos *frames* semânticos que evocam tais unidades.

No *Spanish FrameNet*, a análise semântica de uma unidade léxica se realiza em função do *frame* semântico a qual pertence, partindo das diferentes realizações sintáticas de seus argumentos semânticos, tal como se manifestam. Desde esta proposição, é o *frame* semântico que evoca uma unidade léxica, o que permite a compreensão de seu significado, inclusive quando seus argumentos semânticos não estão presentes na oração. Portanto, a análise semântica que realiza *FrameNet* em relação a uma unidade léxica qualquer consiste, em primeiro lugar, em determinar o *frame* semântico que permite construir seu significado, em função das construções sintáticas governadas por essa unidade léxica tal como aparecem documentadas no corpus.

Quanto ao projeto *FrameCorp*, é uma investigação semântico computacional do léxico do Português do Brasil. Este projeto, único no Brasil, também está baseado na semântica de *Frames* (Fillmore, 1982, 1985), que, a partir do *PLN – BR Gold*, prevê a construção de um corpus com anotação semântica.

Cabe destacar que o *Spanish FrameNet*, bem como o *FrameCorp*, tomam como referência o *FrameNet* do Inglês. Considerando que os *FrameNets* estão todos conectados através de *frames* semânticos, estes possibilitam o comparativo entre as línguas. A anotação dos textos, tanto no *Spanish FrameNet*, quanto nos projetos interligados, está em Língua Inglesa, sendo assim, a investigação adquire um caráter multilíngüe por envolver três idiomas.

Neste projeto, o idioma inglês funcionará como uma interlíngua entre o espanhol e o português, isto é, uma língua de representação independente que contém informação relevante. Ela permitirá fazer abstrações sobre o fenômeno em estudo e realizar comparações. A partir da análise dos casos poderá se verificar o quanto uma língua

pode ser apropriada para descrever classes semânticas de outras línguas. Isto significa que o projeto trabalhará também com conceitos de equivalência e multilinguagem.

3. Objetivos

O objetivo desta pesquisa é estudar os padrões de lexicalização dos verbos psicológicos das línguas espanhola e portuguesa, vertente brasileira, permitindo assim verificar quais os itens lexicais que possuem padrões diversos nestas línguas.

Através da investigação, busca-se responder às seguintes questões:

- a) em que medida a descrição semântica proposta no paradigma *FrameNet*, em Inglês, se presta ao estudo da polissemia no que tange ao contraste Espanhol – Português.
- b) em que medida a descrição semântica proposta pelo paradigma *FrameNet*, em Inglês, se presta à descrição deste fenômeno semântico considerando o contraste Espanhol – Português.

Para poder respondê-las, os objetivos estão divididos em:

Objetivos de análise:

- analisar a polissemia dos verbos psicológicos de Língua Espanhola e Portuguesa, do ponto de vista teórico, verificando em que medida os diferentes padrões de lexicalização podem interferir na anotação semântica de textos em Português a partir das etiquetas semânticas do *FrameNet*.
- investigar o tratamento dado ao fenômeno da polissemia em verbos psicológicos em léxicos computacionais através da Linguística Cognitiva e da Semântica de *Frames*.

Objetivos de aplicação:

- identificar os verbos psicológicos da Língua Espanhola, extraíndo as ocorrências no Espanhol e no Português.
- comparar esses itens lexicais nos dois idiomas, verificando quais apresentam padrões de lexicalização diversos.
- gerar um corpus de léxico bilíngüe usando o corpus disponível através do *Sketch Engine*.

De forma resumida, serão investigadas as diferentes abordagens dadas aos verbos psicológicos por diversos teóricos e o tipo de tratamento dado a estes nos léxicos computacionais. No final do trabalho será proposta uma nova abordagem de categorização para esses verbos buscando um tratamento que dê conta das línguas espanhola e portuguesa. As informações coletadas permitirão verificar o quanto as bases de dados se integram, em que medida são diferentes em cada língua e se há algum frame que dê conta dessa diferença.

4. Metodologia

A investigação será tanto de ordem teórica quanto prática. Na pesquisa teórica será realizada uma investigação através da semântica de *Frames* e da Linguística Cognitiva para verificar qual o tratamento que o fenômeno da polissemia em verbos psicológicos tem recebido nos léxicos computacionais. Na pesquisa aplicada será realizado um estudo sobre o léxico, um recorte dos verbos de natureza psicológica e uma extração das ocorrências no Espanhol e no Português.

Para alcançar os objetivos propostos será realizada uma análise quantitativa, extraído do corpus *Spanish FrameNet*, todos os verbos da língua espanhola que estiverem anotados sob a caracterização de *frame emotion*. Para cada um dos verbos se buscará o seu equivalente na língua portuguesa de variante brasileira e se verificará se este encontra categorizado como verbo pertencente ao *frame emotion*. A análise dos verbos equivalentes em língua portuguesa será realizada no léxico computacional *Sketch Engine*. A partir dos dados coletados poderá se realizar um estudo comparativo buscando responder às questões de pesquisa.

Os resultados de cada língua e do comparativo entre elas serão tabulados e sistematizados estatisticamente através de gráficos detalhados. A partir da análise dos dados, será gerado um corpus de léxico bilíngüe usando o *Sketch Engine* e que servirá como base de informação para o Projeto FrameCorp. Ao final do trabalho, com base nos dados levantados será proposta uma nova abordagem de categorização dos verbos buscando um tratamento que dê conta das línguas espanhola e portuguesa.

De forma bastante resumida, este projeto tem a intenção de poder contribuir de alguma forma com as áreas da Informática e da Linguística, através de um estudo comparativo entre línguas próximas como são o Espanhol e o Português. Também, que o produto originado desta investigação possa vir a ser uma fonte de informação para tradutores, professores, bem como demais interessados no tema.

5. Referências bibliográficas

ALONGE et al. The linguistic design of the EuroWordNet Database. In: Vossen, Piek (ed.) *EuroWordNet: A multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer: Academic Publishers, 1998.

BOAS, H. C. “Bilingual FrameNet Dictionaries for Machine Translation”. In *Proceedings of the Third International Conference on Language Resources and Evaluation*. Las Palmas, Spain. Vol. IV: 1364-1371. 2002.

BOAS, H. C. “Semantic Frames as Interlingual Representations for Multilingual Lexical Databases”. *International Journal of Lexicography*. 2005 18(4): 445-478.

BURCHARDT, A.; ERK, K.; FRANK, A.; KOWALSKI, A.; PADÓ, S. “SALTO -- A Versatile Multi-Level Annotation Tool”. *Proceedings of LREC 2006*, Genoa, Italy. 2006.

CHOMSKY, Noam. **Knowledge of Language**. New York: Praeger, 1986.

CHOMSKY, Noam. Remarks on Nominalization. In: JACOBS, R.; ROSENBAUM, P. (Eds.) **Readings in English Transformational Grammar**, 1970.

COLLINS, A.M., QUILLIAN, M. R. How to Make a Language User. In: TULVING; DONALDSON (Eds.) **Organization of Memory**, 1972.

CRUSE, A. **Lexical Semantics**. Cambridge. Cambridge University Press, 1886.

CRUSE, A. **Meaning in Language: an Introduction to Semantics and Pragmatics**. New York: Oxford University Press, 2000.

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

FILLMORE, C. J. "Frame Semantics". The Linguistic Society of Korea. *Linguistic in the Morning Calm*, Seoul, Hansinh Publishing Co., 111-137. 1982.

FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. "Background to FrameNet". *International Journal of Lexicography*. 16(3):235-250. 2003.

FODOR, J. **Psychosemantics**. Cambridge: MIT Press, 1987.

FONTENELLE, T. (ed.). *International Journal of Lexicography*. Special Issue: FrameNet and Frame Semantics. 16(3):231-385. 2003.

FONTENELLE, T. A Bilingual Lexical Database for Frame Semantics. *International Journal of Lexicography*, 2000; 13(4): 232 - 248.

GARCÍA-MIGUEL, José M. *Las relaciones gramaticales entre predicado y participantes*. Universidad de Santiago de Compostela. (1995)

GOLDBERG, Adele E. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press, 1995.

GRUBER, T. R. A Translation Approach to Portable Ontology Specifications. **Knowledge Acquisition**, 5. p. 109 – 200, 1993.

HASEGAWA, Y.; OHARA, K.; LEE-GOLDMAN, R.; FILLMORE, C. J. "Frame Interegion, Head Switching, and Translation: RISK in English and Japanese." ICCG4. FOURTH INTERNATIONAL CONFERENCE ON CONSTRUCTION GRAMMAR, Tokyo, Japan. September, 2006.

JACKENDOFF, R. **Languages of the Mind**. Cambridge: The MIT Press, 1992.

KATZ, J.; FODOR, J. The structure of a Semantic Theory. In: FODOR, Jerry. KATZ, Jerrold. **The Structure of Language. Readings in the Philosophy of Language**, Englewood Cliffs, New Jersey: Prentice-Hall Inc., p. 479-518, 1963.

LEVIN, Beth. *English Verb Classes and Verbal Alternations. A Preliminary Investigation*, The University of Chicago Press. 1993.

LÖNNEKER-RODMAN, B. Multilinguality and FrameNet. *ICSI Technical Report TR-07-001, Berkeley, CA, March 2007.*

LYONS, J. **Semântica**. Lisboa: Editorial Presença, 1977.

MELIS, C. Variación actual con los verbos de emoción. **Español actual**. 71, 1999, 49-62.

MILLER, G.A., JOHNSON-LAIRD, P. N. **Language and Perception**. Cambridge: The Belknap Press of Harvard University Press, 1976.

MILLER, George. **Dictionaries in Mind, Language and Cognitive Process**. Cambridge: Harvard University Press, 1986.

OLIVEIRA CHISHMAN, R. L. de. Contribuições da Semântica Lexical para a Linguística Computacional: o desenvolvimento de ontologias. In: VIEIRA, Renata., STHRUBE DE LIMA, Vera. **Engenharia da Linguagem**. 2005.

PUSTEJOVSKY, J. **The generative lexicon**. Cambridge: MIT, 1995.

ROVENTINI, A. et al. "Italwordnet: building a large semantic database for the automatic treatment of Italian". In: *Linguistica Computazionale*. Pisa-Roma: Istituti Editoriali e Poligrafici Internazionali, ISSN. p.745-791, 2003.

SAGRI, M. T.; TISCORNIA, D.; BERTAGNA, F. Jur-WorNet. In: SOJKA, P. et al. (Eds.) SECOND INTERNATIONAL WORDNET CONFERENCE - GWC 2004. Brno: Masaryk University, 2003, p.305-310.

SILVA, Augusto Soares. Linguagem, Cultura e Cognição ou A Linguística Cognitiva. In: SILVA et al. (orgs). Linguagem, Cultura e Cognição: Estudos de Linguística Cognitiva. Coimbra: Almedina, vol.I, 2004. p. 1-18.

SILVA, Augusto Soares. Situación Actual de la Lingüística Cognitiva Portuguesa. In: Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística. 2004.

SUBIRATS RÜGGERBERG, C. 2006. FrameNet español. Una red semántica de Frames conceptuales. En E. Serra y G. Wotjak, eds. *Cognición y percepción lingüísticas*. Valencia: Universidad de Valencia y Universidad de Leipzig, págs. 182-196. http://gemini.uab.es/SFN/papers/Leipzig_Paper.pdf

SUBIRATS, C.; PETRUCK, M. Surprise: Spanish FrameNet. *International Congress of Linguists. Workshop on Frame Semantics, Prague (Czech Republic), July 2003.*

SUBIRATS, C.; SAITO, H. 2004. "Spanish FrameNet and FrameSQL". 4TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION. WORKSHOP ON BUILDING LEXICAL RESOURCES FROM SEMANTICALLY ANNOTATED CORPORA, Lisbon (Portugal), May 2004.

USCHOLD, M., KING, M. Towards a Metodology for Buiding Ontologies. In.: Workshop on **Basic Ontological Issues in Knowledge Sharing - IJCAI-95**, Edinburg, 1995.

VOSSSEN, Piek. **Grammatical and conceptual individuation in the lexicon**, Universiteit van Amsterdam, Amsterdam. 1995.