

## CONVERGÊNCIAS ENTRE SEMÂNTICA DE FRAMES E COMPUTAÇÃO

Rove Chishman<sup>1</sup>

---

### RESUMO

Neste trabalho, nosso principal propósito é apresentar a semântica computacional, área que ilustra a interface da semântica com a computação e desempenha um papel central em vários tipos de aplicação para o processamento computacional da linguagem humana. Propomo-nos também a chamar a atenção para o papel que a Semântica de Frames (FILLMORE et al., 2003) vem desempenhando neste cenário, a partir da criação da *FrameNet* e de projetos similares para outras línguas. A título de ilustração, fazemos também um breve relato de duas iniciativas do grupo *SemanTec* que tem a Semântica de Frames como seu principal referencial teórico na tarefa de construção de recursos semântico-computacionais.

**Palavras-chave:** Frames. Semântica computacional. Léxicos computacionais. Informática jurídica.

---

### 1 INTRODUÇÃO

Neste trabalho, nosso principal propósito é apresentar a semântica computacional, área que ilustra a interface da semântica com a computação e desempenha um papel central em vários tipos de aplicação para o processamento computacional da linguagem humana. Propomo-nos também a chamar a atenção para o papel que a Semântica de Frames (FILLMORE et al., 1993) vem desempenhando nesse cenário, a partir da criação da *FrameNet* e de projetos similares para outras línguas. A título de ilustração, fazemos também um breve relato de duas iniciativas do grupo *SemanTec* que tem a Semântica de Frames como seu principal referencial teórico na tarefa de construção de recursos semântico-computacionais. A primeira diz respeito ao estudo dos frames semânticos com vistas à construção de um dicionário trilingue para a linguagem do futebol. A

segunda é voltada para a linguagem jurídica e tem como foco os sistemas de recuperação de informação na web.

## **2 SEMÂNTICA LEXICAL COMPUTACIONAL**

Antes de destacar o desenvolvimento da Semântica Lexical, a partir das demandas da computação, convém chamar a atenção para o desenvolvimento dos estudos do léxico e da semântica no âmbito da história recente da Linguística. Não é demais lembrar dos percalços enfrentados pelas abordagens semânticas que surgiram, no século XX, em face às exigências impostas pela visão de ciência que fundamentou tanto a corrente estruturalista como a gerativista.

Como bem expõe Borges Neto (2003, p. 9), a natureza “não-física” dos significados parece trazer problemas para as teorias que pretendem considerá-los uma espécie de entidade. Teorias semânticas fundamentadas em uma tradição mentalista, ou que seguem uma linha referencial, são igualmente postas em xeque, se o objetivo é garantir à semântica o status de ciência já conquistado por outras áreas, como a sintaxe, por exemplo. A semântica formal ganha espaço nesse cenário ao procurar atender esse quadro exigente, haja vista que se preocupa em dispor de representações semânticas escritas numa linguagem formalmente bem caracterizada. Interessa aos seguidores desse tipo de semântica que a linguagem das representações semânticas seja construída, segundo os princípios que regem a construção das linguagens formais dos lógicos, ou mesmo linguagens já construídas pelos lógicos.

Ainda que pareça meritória e desejável essa tentativa de utilização de linguagens formais como modelo para a linguagem das representações semânticas, são muitas as objeções apontadas a essa visão de semântica lógica. Uma das principais objeções, conforme Borges Neto, é a impossibilidade de as linguagens formais representarem outras funções, como seu aspecto ilocutório, além da função representativa.

Ainda situando os estudos do significado no âmbito de uma linguística formal, temos sinais de reação a esta fase de ostracismo da semântica com a valorização do papel do léxico em diferentes modelos teóricos. Interessa-nos, em especial, fazer referência a duas propostas teóricas que bem ilustram a mudança de paradigma nos estudos semânticos, que trazem repercussões importantes para a relação com a

Computação. As teorias defendidas por Ray Jackendoff (1983, 1990, 2002) e James Pustejovsky (1995) são exemplos de propostas que adotam a denominação lexicalista. Isso, por si só, já é merecedor de nossa atenção, pois é indicativo de que o léxico ocupa uma posição central em seus modelos.

O léxico, nesse sentido, deixa de ser visto como uma mera lista de entradas e passa a ser visto como um rico constructo teórico com sua própria complexidade e propriedades formais. A Semântica Conceptual (JACKENDOFF, 1983) é ilustrativa dessa posição. Concebendo como foco a noção de Conceitos-I (I de Internalizados), Jackendoff propõe-se a caracterizar os recursos mentais que definem uma variedade de conceitos, a partir de seus traços formais. Um dos diferenciais dessa visão é a inclusão de unidades reais do vocabulário e de expressões idiomáticas no componente lexical. Embora sua principal contribuição esteja vinculada à relação da linguagem com a cognição, no campo do processamento da linguagem natural, a Semântica Conceptual tem seu alcance limitado, haja vista que não tem maquinário descritivo para dar conta da representação do conteúdo de categorias, como os nominais, por exemplo. Conforme se discutirá mais adiante, as aplicações computacionais impõem necessidades representacionais que nem sempre são atendidas por completo pelas teorias linguísticas.

A Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995), por sua vez, além de representar uma ruptura com o ceticismo que dominou a semântica nesses anos todos, é delineada tendo a metodologia oferecida pela computação. Sem abrir mão da tratabilidade computacional, Pustejovsky propõe uma agenda ampla de questões semânticas, incluindo temas comumente negligenciados, como a polissemia e a semântica nominal e adjetival. Desenvolvida com o intuito de aproximar Semântica Lexical e Computação, a TLG baseia-se no princípio de que o léxico não pode ser apresentado como um conjunto estático de palavras. Segundo ele, só um modelo mais complexo, formado por um sistema rico de representação do conhecimento lexical, em conjunção com mecanismos de inferência, é capaz de expressar uma série de generalizações linguísticas. A natureza polimórfica da linguagem e o uso criativo das palavras em contextos novos são fenômenos não capturados de forma sistemática por teorias que se valem de técnicas enumerativas de significado.

De acordo com Pustejovsky, há uma série de regularidades linguísticas – independentemente de relações lexicais como sinonímia, antonímia, hiponímia e meronímia – que podem ser geradas pelo sistema, o que resultaria em um léxico

com um número menor de entradas do que o apresentado pela prática comum. O relevante, segundo ele, é considerar a relação lógica entre os sentidos que um item lexical pode assumir como um dos elementos responsáveis pela simplificação das entradas. O sistema deve prever a conexão, por exemplo, entre *janela*, significando objeto físico, e *janela*, significando abertura. São propriedades de paradigmas inteiros que devem ser previstas pelos formalismos.

O que faz desse autor um expoente da Semântica Lexical Computacional é, principalmente, sua preocupação em atender à demanda imposta pelas aplicações e em mesclar os avanços da linguística teórica às novidades tecnológicas em termos metodológicos. A Semântica Lexical representa, nessa visão revigorada, uma área de convergência entre semântica lexical, lexicografia computacional e semântica computacional.

Assim como na linguística teórica, os léxicos computacionais são construídos como dotados de uma grande sistematicidade, possíveis de serem estudados com o objetivo de se obter uma representação compacta e não-redundante. São elaborados com o objetivo de servir de suporte a diversas aplicações, tais como sistemas de tradução automática ou de recuperação e extração de informação na web. É nesse sentido que os estudos em semântica lexical têm seu escopo ampliado: cabe à semântica linguística subsidiar esses estudos aplicados, fornecendo representações lexicais mais elaboradas e mais estruturadas sobre amplas áreas do vocabulário, ou seja, descrições lexicais que tenham fundamentação teórica.

A importância de se ter um componente lexical semanticamente enriquecido se deve, na visão de Saint-Dizier e Viegas (1995), ao fato de as entradas lexicais conterem uma parte considerável de informação, relacionada ao sentido que representam. A construção de bases de dados lexicais e ferramentas de construção de léxicos computacionais pode, nesse sentido, se beneficiar de uma caracterização detalhada do significado das palavras, haja vista que passa a incluir informações estruturadas que podem melhorar a qualidade da aplicação.

Compartilhamos com Boguraev e Levin (1993, p. 326) a premissa de que uma teoria linguística é central para a construção de léxicos completos e bem fundamentados. Ainda que trabalhosa, uma análise linguística cuidadosa parece desejável, pois fornece dados mais confiáveis. A perspectiva que se quer focalizar, então, privilegia a integração de estudos teóricos em semântica lexical e o trabalho

sobre representação do conhecimento e design do léxico em computação. Em outras palavras, estão em jogo léxicos semânticos tratáveis computacionalmente, mas teoricamente motivados.

### 3 SEMÂNTICA LEXICAL: UMA ORIENTAÇÃO A PARTIR DA SINTAXE

Há diferentes tradições no estudo da semântica lexical. Duas delas, em especial, parecem ser comumente empregadas nas aplicações computacionais. Uma baseia-se nos estudos sintáticos e a outra está relacionada às abordagens originárias da Inteligência Artificial. É comum também fazer referência às duas visões como contrárias. A primeira objetiva descobrir propriedades semânticas dos itens lexicais que repercutem diretamente na sintaxe. Denominações como semântica gramatical ou semântica linguística referem-se a esta primeira visão. Nirenburg e Levin (1991) referem-se a esta perspectiva, empregando a denominação *syntax-driven lexical semantics* ou *supply-side semantics*. Entre as vantagens de se assumir este tipo de abordagem, é poder representar os sentidos de predicados e argumentos, expressar o tipo de dependência entre núcleos e argumentos e mapeá-los à sintaxe. Em um ambiente computacional, ajuda a distinguir os significados, embora só represente uma parcela do conteúdo.

Informações sobre os diferentes papéis temáticos que podem vir a ocupar uma posição sintática podem ser incluídas nas entradas lexicais. Entre as peculiaridades da visão sintaticamente orientada, destaca-se o interesse em formular teorias – em semântica lexical – que focalizem a atenção na elegância formal, na economia de recursos descritivos e na ausência de exceções. Em consequência desse tipo de compromisso, fenômenos que não se submetem ao rigor almejado, não são tratados. É o que normalmente ocorre com as questões de conteúdo lexical, que raramente são discutidas por teorias que seguem essa tradição. Estuda-se o que é possível.

Como exemplo dessa concepção, temos os estudos de Grimshaw e Levin no âmbito do *MIT Lexicon Project*. Relevantes para esses estudos é o mapeamento entre léxico e sintaxe, investigando como a informação lexical interage com as propriedades sintáticas. Destacam-se os estudos sobre alternâncias verbais, como em *Maria abriu a porta* e *A porta abriu*, que passaram a ser explicadas a partir da noção de papéis temáticos.

A Semântica Conceptual proposta por Jackendoff também ilustra essa visão. As chamadas Estruturas Conceptuais Lexicais (*LCSs*, de *Lexical Conceptual Structures*) e as Regras de Ligação são os elementos que constituem o conhecimento lexical. As *LCSs* de uma determinada língua são constituídas de um conjunto universal de blocos construtores. A propriedade mais importante das *LCSs* é conter a informação necessária para as Regras de Ligação calcularem a codificação sintática de predicados e argumentos. As Regras de Ligação descrevem como as *LCSs* são sintaticamente representadas em uma língua e dividem as *LCSs* em classes de predicados que têm codificação sintática semelhante. Trata-se, pois, de informação semântica sintaticamente relevante.

Relacionada à proposta de Jackendoff está a solução apresentada por Dorr (1995) para tratar dos problemas de variações – as chamadas divergências – enfrentados pela área da tradução automática. A autora, adotando uma visão de interlíngua para esta aplicação, propõe a construção de uma representação independente de línguas particulares, uma espécie de representação semântica subjacente para a tradução que permita o mapeamento entre a representação interna do sistema e a estrutura sintática, acomodando as divergências.

Entre as divergências estudadas por Dorr, temos a questão temática. Comparando o Inglês com o Espanhol, percebe-se que o tema é realizado como o complemento verbal em Inglês e como sujeito do verbo principal em Espanhol, como em *I like Mary* (I) e *María me gusta* (E). Com base nas estruturas conceptuais lexicais propostas por Jackendoff, Dorr sugere o uso de “representações semânticas subjacentes” e interessa-se pela identificação de universais para a tradução.

O sistema de tradução automática UNITRAN (DORR, 1987), que traduz Inglês, Espanhol e Alemão de forma bidirecional, baseia-se em um design que prevê uma classificação para os casos de divergência e implementa um mapeamento sistemático entre a interlíngua semântico-lexical e a estrutura sintática superficial, que apresenta as divergências. Como resultado, é possível isolar os componentes de significado que são comuns aos verbos que fazem parte de classes semânticas específicas. Esses componentes podem ser usados para determinar a representação de verbos, que independe das línguas particulares.

Citamos a área da Tradução Automática e a aplicação que Dorr propõe, a partir da Semântica Conceptual de Jackendoff, para ilustrar o emprego de uma abordagem semântica calcada na sintaxe em uma aplicação voltada para o PLN.

Parece evidente que a incorporação de informação estrutural para a realização e posicionamento de argumentos é parte importante para se garantir o sucesso da TA.; contudo, não pretendemos aqui fazer uma apreciação aprofundada das reais vantagens e limitações de se adotar tal concepção teórica.

Para esses pesquisadores, o interessante está na possibilidade de identificação de regularidades, o que indicaria a existência de princípios gerais de organização lexical, sejam universais ou específicos a línguas particulares.

#### **4 SEMÂNTICA LEXICAL: UMA ORIENTAÇÃO A PARTIR DA DEMANDA COMPUTACIONAL**

Analisando os léxicos computacionais sob a perspectiva semântica, deparamo-nos com abordagens já delineadas, tendo a aplicação em mente, visão diferente da apresentada anteriormente. Niremburg e Raskin (2004) denominam esta visão de *demand-side lexical semantics*, em oposição à visão anterior, as abordagens de orientação sintática, as *supply-side lexical semantics*, também chamadas de *ontology-driven lexical semantics*. Enquanto a primeira escola, a de orientação sintática, põe o foco nos componentes do significado lexical que determinam o comportamento sintático, a tarefa das abordagens que seguem essa segunda concepção consiste em determinar e especificar significados de unidades lexicais, além de sugerir como as unidades podem contribuir para a representação geral do significado do texto. Trata-se, pois, de uma tarefa vinculada a uma proposta mais ampla de representar o significado de um texto em uma aplicação computacional. Esta visão tem sido desenvolvida especialmente na tradição de PLN.

Niremburg e Raskin destacam que tal concepção parte de uma atitude “construtiva”, se comparada à semântica dos modelos. Uma das formas de construir uma ontologia, afirmam esses autores, é propor um conjunto de propriedades que descrevam coisas no mundo, definir conjuntos de valores para essas propriedades e, por conseguinte, descrever cada conceito no mundo como um conjunto de pares de valores particulares.

Em um ambiente computacional, a semântica lexical sintaticamente orientada parece ajudar a distinguir significados, mas não os representa. Isso parece indicar que, em casos de ambiguidade, a análise predicado-argumento não é suficiente. Há necessidade de representar o significado para se escolher o sentido.

Comparando as duas concepções, pode-se dizer que a primeira só se ocupa dos fenômenos possíveis, enquanto a segunda escola, baseada nas necessidades de descrição das aplicações computacionais, focaliza o que deve ser feito para garantir o valor de uma aplicação. Em face a esse compromisso com a aplicação, trata-se de concepções teóricas que enfatizam as questões básicas de significado, de maneira a não dissociá-lo do conhecimento de mundo. A profundidade e a cobertura da descrição do léxico dependem da arquitetura do sistema em desenvolvimento. Para tarefas que exigem interpretação rigorosa, a escolha da representação deve ser feita com mais cuidado. O importante é construir modelos (isto é, ontologias) e pô-los em correspondência com o léxico, que passaria a conter conhecimento que não é exigido para a decodificação da realização sintática dos argumentos.

Nirenburg e Raskin (2004), referindo-se a essa perspectiva como uma das premissas fundamentais para a realização de uma descrição semântica para compor um léxico computacional, insistem na necessidade de se considerarem, para fins de descrição semântica, os casos ordinários. Segundo eles, a linguística teórica têm se debruçado sobre fenômenos que são excepcionais na língua. Em vez disso, defendem que se desenvolva uma sociedade de microteorias que possam dar conta do fenômeno da linguagem como um todo. Entre os argumentos favoráveis a essa concepção de microteorias, está a facilidade de congregar contribuições da linguística e da linguística computacional, em se tratando de um fenômeno em particular. O problema em se seguirem as descrições estritamente linguísticas é estarmos lidando com fragmentos do conjunto mais amplo do fenômeno linguístico. Para fins de aplicação computacional, a relação entre linguística teórica e linguística computacional deve ser baseada no conceito de uma “teoria da prática”, o que significa se valer de uma teoria para aplicar as descobertas da linguística teórica e descritiva no PLN. Dessa forma, evita-se a tentativa de transpor os resultados da linguística diretamente às aplicações computacionais, sem considerar as diferenças de premissas e objetivos ou, alternativamente, a tentativa de ignorar descobertas teóricas já consolidadas.

O projeto MikroKosmos (NIRENBURG et al., 1996, cujo objetivo é desenvolver um componente de um sistema de tradução automática, baseado em conhecimento, parte do conceito de microteoria para a descrição semântica das categorias. Em se tratando, por exemplo, da análise proposta para os adjetivos, a



microteoria do MikroKosmos indica que as taxonomias e as questões que são importantes para a representação e determinação do significado não o são na literatura. Em outras palavras, o que pode ser essencial em um trabalho de descrição em um léxico computacional não é descrito como tal na literatura.

Entre os expoentes dessa forma de abordar a semântica lexical, destacam-se Sowa (2000), Schank (1973) e Wilks (1972). A Teoria do Léxico Gerativo, de James Pustejovsky, é um bom exemplo de como é possível reunir as duas concepções: há uma preocupação com uma semântica de eventos, que, por sua vez, está associada a uma estrutura de argumentos, e há também espaço para a veiculação do conteúdo dos nominais, no nível conhecido de estrutura qualia, característica dessa segunda visão apresentada.

Levin (1995) também é favorável a essa solução híbrida. Uma abordagem compreensiva deve combinar as vantagens de ambas as visões: a visão de orientação sintática incluiria um componente de conhecimento lexical, descrevendo a codificação sintática, o que auxiliaria a distinguir significados; a visão de orientação ontológica contribuiria com a força expressiva adicional, ao fornecer informação residual à estrutura predicado-argumento, como, por exemplo, especificações quanto à restrição seletional.

## **5 A SEMÂNTICA DE FRAMES COMO UM MODELO PARA O PLN**

É seguindo o conceito de microteoria e esta visão híbrida proposta por Niremburg e Levin que elegemos a Semântica de Frames e sua contraparte computacional, a *FrameNet*, como modelos centrais para a construção de léxicos computacionais e ontologias. Entre os principais diferenciais dessa abordagem, ao se ter em mente a representação do conhecimento semântico em modelos computacionais, é sua capacidade descritiva, em especial, quando se pensa em papéis situacionais. Nesse sentido, a noção de *frame* e os papéis semânticos parecem ser desejáveis, se comparados aos papéis temáticos, na medida em que permitem uma descrição rica para diferentes cenários.

Importa destacar que essa abordagem não foi delineada em sua origem, tendo em vista as demandas computacionais, como é o caso da TLG de Pustejovsky.

A Semântica de *Frames* nasce a partir de um conceito muito discutido na década de 70, o *frame* (MINSKY, 1974; GOFFMAN, 1974). Segundo Fillmore (1982), o *frame* representa um sistema de conceitos relacionados entre si, em que o entendimento de um dos conceitos desta rede de conceitos implica a compreensão dos demais conceitos. Para a Semântica de *Frames*, as palavras têm capacidade de “evocar” todo um conhecimento de mundo que é organizado através de uma estrutura cognitiva chamada *frame* (FILLMORE, 1985).

Conforme Boas (2009, p. 68), o *frame* semântico se refere a um “plano de fundo comum de conhecimento, no qual o significado das palavras passa a ser interpretado”. Dessa forma, é possível compreender significados, a partir de estruturas de conhecimento que são socialmente compartilhadas, graças às situações vividas e compreendidas pelos envolvidos. Fillmore (1982) caracteriza-o como um sistema de conceitos relacionados, no qual, para compreender um deles, é necessário compreender o funcionamento do todo, ou seja, “quando uma das coisas que faz parte da estrutura é apresentada em um texto, ou em uma conversa, todas as demais envolvidas são automaticamente relacionadas” (FILLMORE, 1982, p. 111). Assim, as palavras estão relacionadas às nossas experiências e aos conhecimentos prévios e não são compreendidas numa visão de dicionário, mas numa visão enciclopédica, como defende a Linguística Cognitiva.

Coube ao projeto *FrameNet* (FILLMORE; JOHNSON; PETRUCK, 2003) o desenvolvimento da contrapartida computacional da Semântica de *Frames*. Em forma de base de dados, a *FrameNet* é um recurso lexical monolíngue que descreve o significado das unidades lexicais da língua inglesa, com relação aos *frames* semânticos que cada unidade lexical evoca. No entanto, o desenvolvimento de projetos semelhantes, como a *German FrameNet* (BOAS, 2002; 2005), a *Spanish FrameNet* (SUBIRATS; PETRUCK, 2003) e a *Japanese FrameNet* (OHARA et al., 2003), estimulou a discussão sobre o uso da *FrameNet* como ponto de conexão entre bases de dados lexicais de línguas diferentes, criando um léxico multilíngue (BOAS, 2005).

Além dos recursos lexicais já citados, estão em construção recursos lexicais baseados em *frames* para o hebraico (PETRUCK, 2009), o português brasileiro (SALOMÃO, 2009), o italiano (VENTURI et al., 2009) e o francês (PITEL, 2009). Os *frames* e elementos de *frame* da *FrameNet* também têm sido utilizados para anotação de *corpora* em alemão (BURCHARDT et al., 2009). Todos os recursos

lexicais até aqui citados estão voltados à descrição de língua geral. No entanto, a *FrameNet* também serviu de inspiração para o desenvolvimento de recursos léxico-computacionais para linguagens especializadas. Um exemplo de aplicação dos *frames* para linguagens especializadas é o Kicktionary (SCHMIDT, 2009), base de dados lexicais multilíngue da linguagem do futebol e fonte de inspiração para o projeto Kicktionary-Br (CHISHMAN, 2010).

## 6 O PROJETO KICKTIONARY-BR COPA 2014

Os estudos realizados no escopo do projeto Kicktionary ilustram a aplicação da Semântica de Frames, em duas frentes da Linguística Computacional: a anotação semântica de corpus e a lexicografia computacional.

O trabalho de anotação semântica de corpus consistiu em aplicar as categorias propostas por Schmitt (2009) na etiquetagem de um corpus constituído de 3000 textos. Destacam-se aqui a metodologia empregada de anotação de texto corrido (*running text*) e, como decorrência desta metodologia, o enfrentamento de fenômenos linguísticos específicos, como é o caso da alta incidência de metáforas, construções com verbos suporte e colocações, todos eles desafios, no que tange à descrição baseada em frames. Outro aspecto importante deste trabalho foi a necessidade de contar com um editor de anotações, de modo que o corpus anotado ficasse em linguagem legível por máquina. Optamos pela ferramenta SALTO, especialmente desenvolvida para anotação de frames.

Como desdobramento do projeto de anotação de corpus, surge o projeto Kicktionary-Br Copa 2014<sup>2</sup>, que tem como principal objetivo teórico refletir sobre a pertinência da abordagem baseada em frames, para a construção de um recurso lexicográfico trilingue (português-inglês-espanhol). Em termos aplicados, nosso propósito é desenvolver um aplicativo, tendo em vista a demanda imposta pelo evento da Copa do Mundo 2014.

Diferentemente do trabalho de anotação, o dicionário da Copa nos impôs investigar os preceitos da lexicografia, incluindo as várias etapas que constituem a construção de um dicionário, como (i) definição do perfil do usuário; (ii) construção de um guia de estilo; (iii) planejamento da micro, macro e megaestrutura; (iv) definição de estrutura de acesso e de referências cruzadas e (iv) estudo de equivalentes de tradução.

## 7 O PROJETO CNJ – ACADÊMICO E A INFORMÁTICA JURÍDICA

O projeto CNJ-Acadêmico<sup>3</sup> parte do estudo da Semântica de Frames, tendo como propósito sua aplicação a uma outra área importante do PLN: os sistemas de recuperação de informação, mais especificamente os sistemas de informática jurídica. De caráter multidisciplinar – envolvendo a *expertise* das áreas da Linguística, com ênfase na Semântica Computacional, da Computação e do Direito, o grupo vem se dedicando ao desenvolvimento e à implementação de um modelo semântico-conceitual do domínio jurídico brasileiro, de modo a ser integrado a sistemas de busca e recuperação de informação em *sites* que armazenam documentação jurídica. A opção por utilizar a Semântica de *Frames* (FILLMORE, 1982) como paradigma teórico, para o desenvolvimento de um recurso léxico-computacional especializado, e não um modelo de relações léxico-conceituais, como o paradigma teórico que norteia as *wordnets*, deve-se à possibilidade de mapeamento dos termos relacionados a um determinado evento jurídico que se está descrevendo. Conforme os princípios teóricos da Semântica de *Frames*, entender o significado de uma palavra requer o conhecimento de todos os conceitos que estão relacionados a essa palavra. Em outras palavras, entender o significado de *julgar* requer do falante o conhecimento de outros conceitos que estão relacionados ao verbo *julgar*, como *juiz*, *réu*, *delito* e *sentença*. Assim, a Semântica de *Frames* permite a relação dos participantes e dos eventos jurídicos por meio dos *frames* semânticos.

## 8 CONSIDERAÇÕES FINAIS

Nosso objetivo, neste trabalho, foi mostrar em que medida a teoria da semântica de frames converge com os interesses da Linguística Computacional. Um dos aspectos que procuramos destacar foi a forma diferenciada como se dá o trabalho nesta área, na medida em que se assume uma metodologia híbrida, combinando abordagens aparentemente muito distintas, com vistas a resolver determinada demanda imposta pela computação.

Esta posição serve de base para as duas pesquisas aqui ilustradas: uma diz respeito à anotação semântica de corpus e à construção de dicionários, e a outra, à informática jurídica. Ainda que se debrucem sobre linguagens bem distintas (a do

futebol e a jurídica), as duas investigações ilustram a pertinência da semântica de frames para a Linguística Computacional, haja vista seu potencial descritivo e ontológico.

Um linguista pouco familiarizado com a área de PLN pode não aceitar a metodologia apresentada (*demand side*), o que explica muitas das dificuldades de se trabalhar de forma interdisciplinar. É inegável, contudo, que foi graças às necessidades computacionais que se alavancaram os estudos em semântica.

Gostaria de finalizar chamando a atenção para os percalços de se apostar numa atividade envolvendo competências distintas, como é próprio de qualquer atividade interdisciplinar. O linguista ainda corre o risco de não ter seu trabalho reconhecido nem pelos linguistas teóricos, que ainda não despertaram para este trabalho na interface, nem pelos informatas, que consomem as teorias linguísticas sem conhecê-las. Esta tensão ocorre porque esta área ainda não tem um espaço estabelecido nos cursos de graduação e pós-graduação. Ainda que haja riscos, vale sinalizar para o vasto mercado que se abre para o profissional de Letras, que passa a ter a possibilidade de atuar em atividades que não sejam só de ensino.

## NOTAS

<sup>1</sup> Professora e Pesquisadora no Programa de Pós-Graduação em Linguística Aplicada na UNISINOS. É pesquisadora do CNPq (Bolsa de Produtividade Nível 2. Coordena o grupo de pesquisa SemanTEC (credenciado pelo CNPq).

<sup>2</sup> Projeto desenvolvido em parceria com a Framenet Brasil, sob coordenação da Professora Doutora Margarida Salomão na Universidade Federal de Juiz de Fora.

<sup>3</sup> O projeto CNJ-Acadêmico, intitulado Tecnologias Semânticas aplicadas a Sistemas de Recuperação de Informação Jurídica, recebeu recursos através do Edital CAPES/CNJ Acadêmico 2010.

---

## CONVERGENCES BETWEEN FRAME SEMANTICS AND COMPUTING

### ABSTRACT

In this paper, our main purpose is to introduce computational semantics, area that illustrates the interface of semantics with computing and plays a central role in many types of applications for computational processing of human language. We propose

also to draw attention to the role Frame Semantics (FILLMORE et al., 1993) has played in this scenario, since the creation of FrameNet and similar projects for other languages. As an illustration, we also do a brief account of two initiatives of the SemanTec group, which has Frame Semantics as its main theoretical framework on the task of constructing semantic-computing resources.

**Keywords:** Frames. Computational Semantics. Computational Lexicons. Legal Informatics.

## REFERÊNCIAS

ATKINS, B.T. Sue; RUNDELL, Michael. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press. 2008.

AULETE DICIONÁRIO. Disponível em: <<http://aulete.uol.com.br>>. Acesso em: 3 ago. 2012.

BOAS, Hans C. Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*. [s.l.], 18(4), p. 445-478, 2005.

\_\_\_\_\_. *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin: Mouton de Gruyter, 2009.

\_\_\_\_\_. Bilingual FrameNet dictionaries for machine translation. *Proceedings of the Third International Conference on Language Resources and Evaluation*, Las Palmas, v. IV, p. 1364-1371, 2002.

BOGURAEV, Branimir; LEVIN, Beth. Models for lexical knowledge bases. In: PUSTEJOVSKY, James. (Ed.). *Semantic and the lexicon*. Dordrecht, Nederland: Kluwer Academic Publishers, 1993.

BORGES NETO, José. Semântica dos modelos. In: MÜLLER, A. L.; NEGRÃO, E.; FOLTRAN, M. J. (Orgs.). *Semântica formal*. São Paulo: Contexto, 2003.

BURCHARDT, A. et al. Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In: BOAS, H. C. (Ed.). *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin; New York: Mouton de Gruyter, 2009. p. 209-244.

CHISHMAN, Rove Luiza de Oliveira; FIGUEIRÓ, Guilherme Tiecher. *Anotação semântica de frames dos verbos de movimento na temática do futebol*. In: IX ENCONTRO DO CÍRCULO DE ESTUDOS LINGUÍSTICOS DO SUL – CELSUL,

2010, Palhoça (SC). *Anais...* Palhoça: Universidade do Sul de Santa Catarina, 2010.

DORR, Bonnie. UNITRAN: an interlingual approach to machine translation. *Proceeding of the Sixth Conference of the American Association of Artificial Intelligence*, Seattle, Washington, p. 534-539, 1987.

\_\_\_\_\_. A lexical-semantic solution to the divergence problem in machine translation. In: SAINT-DIZIER, P.; VIEGAS, E. (Eds.). *Computational lexical semantics*. Cambridge University Press, 1995.

FILLMORE, C. J. Frame semantics. *Linguistics in the morning calm*, Seoul, Korea, p. 111-138, 1982.

\_\_\_\_\_. *Frames and the semantics of understanding*. [s.l.], v. 6, n. 2, p. 222-254, 1985.

\_\_\_\_\_. Frame semantics. In: GEERAERTS, D. (Ed.). *Cognitive linguistics: basic readings*. Berlin: Mouton de Gruyter, p. 373-400, 2006.

FILLMORE, C. J.; JONHSON, C.; PETRUCK, M. Background to FrameNet. *International Journal of Lexicography*, v. 16 n. 3, p. 235-250, 2003.

GOFFMAN, Ervin. *Frame analysis*. New York: Harper, 1974.

JACKENDOFF, Ray. *Semantics and cognition*. Cambridge, Mass.: MIT Press, 1983.

\_\_\_\_\_. *Semantic structures*. Cambridge, Mass.: MIT Press, 1990.

\_\_\_\_\_. *Foundations of language*. Oxford: Oxford University Press, 2002.

LEVIN, Beth. Building a lexicon: the contribution of linguistic theory. *International Journal of Lexicography*, v. 4, p. 205-226, 1991.

LORENTE, Mercè. A Lexicologia como ponto de encontro entre a gramática e a semântica. In: ISQUERDO, A., KRIEGER, M. da G. (Orgs.). *As ciências do léxico: lexicologia, lexicografia, terminologia*, v. II. Campo Grande: Editora da UFMS, 2004.

MINSKY, Marvin. A framework for representing knowledge. *Artificial Intelligence*, 1974.

NIRENBURG, Sergei ; LEVIN, Lori. Syntax-driven and ontology-driven lexical semantics. In: PUSTEJOVSKY, J.; BERGLER, S. *Lexical semantics and knowledge representation*. Proceedings of a Workshop. Berkeley: ACL, 1991.

NIREMBURG, Sergei, RASKIN, Victor. Ontological semantics: formal ontology and ambiguity. *FOIS '01*, p.151-161, 2001.

\_\_\_\_\_. *Ontological semantics*. Cambridge, Mass.: MIT Press, 2004.

NIRENBURG, Sergei et al. Lexicon in the Mikrokosmos Project. *Proceedings of the AISB'96 Workshop on Multilinguality in the Lexicon*, Brighton, UK, abr. 1996.

OHARA, Kyoko; HIROAKI, Sato, Fujii; ISHZAKI, Shun. The Japanese FrameNet Project: a preliminary report. *Proceedings of Pacific Association for Computational Linguistics*, Halifax, Canada, p. 249-254, aug. 2003.

PETRUCK, Miriam R. L. Frame semantics. In: ÖSTMAN, J-O; VERSCHUEREN, J.; BLOMMAERT, J. (Eds.). *Handbook of pragmatics*. Amsterdam; Philadelphia: John Benjamins, 1996.

\_\_\_\_\_. Typological considerations in constructing a Hebrew FrameNet. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin; New York: Mouton de Gruyter, 2009. p. 163-182.

PITEL, G. Cross-lingual labeling of semantic predicates and roles: a low-resource method based on bilingual L(atent) S(emantic) A(nalysis). In: BOAS, H. C. (Ed.). *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin; New York: Mouton de Gruyter, 2009. p.245-284.

PUSTEJOVSKY, James. *The generative lexicon*. Cambridge, Mass.: MIT Press, 1995.

SAINT-DIZIER, Patrick; VIEGAS, Evelyne. *Computational lexical semantics*. Cambridge University Press, 1995.

SALOMÃO, Maria M. M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*, v. 7.3, 2009.

SCHANK, R. *Computer models of thought and language*. New York: W. H. Freeman Publisher, 1973.

SCHMIDT, Thomas. The Kicktionary: a multilingual lexical resource of football language. In: BOAS, Hans C. *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin: Mouton de Gruyter: 2009, p. 101-132.

SOWA, John. *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove, CA: Brooks Cole Publishing Co., 2000.

SUBIRATS, Carlos; PETRUCK, Miriam. Surprise: Spanish FrameNet! *International Congress of of Linguists. Workshop on Frame Semantics*, Praga, jul. 2003.

VENTURI, G.; et al. Towards a FrameNet resource for the legal domain. *Proceedings of the Third Workshop on Legal Ontologies and Artificial Intelligence Techniques, Workshop on Semantic Processing of Legal Texts*, Barcelona, jun. 2009.

WILKS, Yorick. *Grammar, meaning and the machine analysis of language*. London; Boston: Routledge. 1972.