



Calidoscópico

E-ISSN: 2177-6202

calidoscopio@unisinis.br

Universidade do Vale do Rio dos Sinos  
Brasil

Martins Salomão, Maria Margarida  
FrameNet Brasil: um trabalho em progresso  
Calidoscópico, vol. 7, núm. 3, septiembre-diciembre, 2009, pp. 171-182  
Universidade do Vale do Rio dos Sinos

Disponível em: <http://www.redalyc.org/articulo.oa?id=571561888002>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica  
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Maria Margarida Martins Salomão  
mm.salomao@uol.com.br

# FrameNet Brasil: um trabalho em progresso<sup>1</sup>

## FrameNet Brasil: A work in progress

**RESUMO** – Apresentação do projeto de anotação lexicográfica FrameNet Brasil, sua base teórica (a semântica de frames), sua vinculação intelectual e metodológica ao projeto FrameNet, de Berkeley, na Califórnia, sua presente etapa de implementação e as reflexões que esta suscita sobre a semântica linguística.

**Palavras-chave:** semântica lexical, frames, léxico, lexicografia.

**ABSTRACT** – Presentation of the project of lexicographical annotation FrameNet Brasil, which includes its theoretical support (frame semantics), its connection to the project FrameNet, in Berkeley, California, its present stage of implementation, and the issues that this kind of semantic analysis suggests about meaning.

**Key words:** lexical semantics, frames, lexicon, lexicography.

### Introdução

Neste artigo, apresentamos o Projeto FrameNet Brasil, de pesquisa lexicográfica, em desenvolvimento na Universidade Federal de Juiz de Fora, com o apoio da FAPEMIG (APQ-01021-08). Na primeira seção, identificamos e caracterizamos sua base teórica, a semântica de frames; a seguir, explicitamos sua vinculação intelectual e metodológica ao Projeto FrameNet, liderado pelo Professor Charles Fillmore no International Computer Science Institute (ICSI), em Berkeley, na Califórnia. As seções subsequentes descreverão as condições em que a FrameNet Brasil vem se implantando – os corpora que constituem sua base de dados, seus primeiros objetos de análise, os resultados preliminares que vai obtendo, e, principalmente, as reflexões suscitadas por esse esforço analítico sobre a semântica linguística.

### A semântica de frames

A semântica de frames é esforço intelectual que retrocede ao final da década de setenta do século passado, quando duas condições contextuais favoreceram o seu florescimento:

- (a) Em primeiro lugar, a atenção que o conceito de frame passou a merecer em várias disciplinas circunvizinhas à linguística: um exemplo bem conhecido na linguística textual praticada no

Brasil, o *script de restaurante*, procede de estudos em Inteligência Artificial (Schank e Abelson, 1977); os estudos sociológicos da interação, desenvolvidos por Erving Goffmann, tematizam a noção de frame como a categoria que coesiona e investe de sentido uma atividade comunicativa (dois meninos engalfinhados no chão podem estar *brigando* ou podem estar *brincando*) (Goffmann, 1974); George Lakoff, ao discutir a radialidade das categorias cognitivas, postula para o conceito de *mãe* diversos Modelos Cognitivos Idealizados (IMCs), que se entrecruzam (Lakoff, 1987); outros muitos exemplos poderiam ser enumerados provenientes do campo dos estudos cognitivos.

- (b) Junte-se a isso, o interesse que levou vários linguistas, originários do empreendimento gerativista, mas dele dissidentes, a buscar soluções para a semântica linguística na continuidade postulada entre a linguagem e as demais capacidades cognitivas (Fillmore, 1976, 1977; Lakoff, 1977; Langacker e Munro, 1975; Langacker, 1982).

A noção de frame emerge na semântica linguística, em artigo de Fillmore (1975), dedicado a demonstrar as insuficiências do tratamento componencial da significação lexical – ou seja, as insuficiências da descrição do significado lexical em termos de uma lista de condições necessárias

<sup>1</sup> Este artigo contou para sua elaboração com a contribuição das Mestrandas Renata Cristina de Barros Vieira Marques, Gabriela da Silva Pires e Paula Zagotta de Oliveira, do Bolsista de Apoio Profissional William Ducca Fernandes e da Bolsista de Iniciação Científica Ludmila Meireles Lage.

(os “componentes”, ou “traços semânticos”), cuja junção constituiria a condição suficiente para a pertinência de um lexema a uma dada categoria significativa.

O caso em disputa nesse artigo correspondia à descrição, por Katz, da polissemia do lexema *bachelor*, do Inglês: adotando-se uma solução maximamente generalista (segundo a piada de Roman Jakobson, o “arquissemema” de *bachelor* corresponderia a {indivíduo do sexo masculino com uma deficiência temporária na realização de uma determinada função social ou biológica}), ou adotando-se uma solução maximamente particularizante, em que se representasse como um feixe de traços cada uma das acepções dicionarizadas do lexema, o fato é que qualquer alternativa geraria resultados deficientes se se buscasse manter, a qualquer preço, a dicotomia, tão cara aos estudos clássicos de semântica, entre dicionário e enciclopédia. Argumenta Fillmore que, considerada a definição em traços de uma das acepções de *bachelor* como {HOMEM & MADURO & NÃO CASADO}, certamente o Papa poderia ser incluído como um exemplar dessa categoria: se essa hipótese causa riso, o despropósito que ela implica decorre exatamente do fato de que a suposta exemplificação desconsidera que o Papa, como indivíduo, ao assumir uma ordem religiosa (a saber, o sacerdócio católico), torna socialmente irrelevante a condição NÃO CASADO da prévia definição. Como operar, então, a aplicabilidade de definições como esta caso a caso?

Fillmore sustenta que definições desse tipo são irre recuperáveis e o são precisamente porque presumem que as significações lexicais tenham caráter discreto e atomístico. O absurdo nas aplicações da definição de *bachelor* resulta do fato de que os conceitos evocados por esse lexema são, na verdade, modelos culturais imbricados em outros modelos culturais: dizer que alguém é NÃO CASADO requer uma definição de CASAMENTO; de outro lado, para que o CASAMENTO como relação entre duas pessoas possa se estabelecer é necessário garantir a legitimidade da pré-condição de NOIVO; ora, quem assume ordens religiosas, dentro do modelo cultural do SACERDÓCIO CATÓLICO, está impedido de satisfazer essa pré-condição. Requer-se, então, para dar conta de todos os fatos, que a noção de significação lexical se torne internamente estruturada e complexa, de tal modo a motivar uma escala pragmática de aceitabilidade dos usos correntes: será talvez possível qualificar o Papa como *bachelor*, desde que se ressalve que o Papa não é um *bachelor* típico...

Discussões desse tipo ensejam a enunciação do grande princípio norteador da semântica de frames: *significações são relativizadas a cenas* (Fillmore 1977), expressando a continuidade básica entre linguagem e experiência. Nesses termos, postula-se a *categoria teórica de frame como estrutura conceptual complexa, organizada de tal modo que, para compreender qualquer de suas partes, é imprescindível o conhecimento do todo*. Expressões

linguísticas evocam frames para serem significativas, embora frames constituam esquemas conceptuais não exclusivamente linguísticos. Ao evocarem um frame, as expressões linguísticas impõem sobre ele uma perspectiva determinada: como no exemplo de Langacker (1987, vol. 1), é impossível compreender a significação do lexema *hipotenusa* sem evocar o frame visual de TRIÂNGULO RETÂNGULO; *hipotenusa*, porém, perfila esse frame de modo diverso ao modo como o perfila o lexema *cateto*. A ideia-chave é que toda unidade lexical evoca um frame, mas, ao fazê-lo, perfila (ou destaca) algum elemento desse frame de forma particular.

A conexão originária entre frame e modelo cultural é fácil de ser reconhecida. Para ficarmos no frame do CASAMENTO (na sua versão ocidental), é impossível ignorar sua relevância para definir ESTADOS CIVIS tais como os designados pelos lexemas *companheira*, *concubina*, *viúva*, *separada* (que, no Brasil, contrasta juridicamente com *divorciada*), ou para definir o que se entende por *pensão alimentícia*. Nesses casos, é bastante óbvio que o “conhecimento do mundo” patrocina de forma organizada e previsível toda a atividade inferencial. Interpretar, por exemplo, uma sentença como *Tem um camelô na esquina vendendo cds genéricos*, requer que sejamos capazes, além de evocar o frame de COMÉRCIO, de mobilizar sua instanciização brasileira na forma de COMÉRCIO AMBULANTE, e, afora isso, compreender o eufemismo que substitui *genéricos* por *piratas*, cujo entendimento requer um outro frame: o de FRAUDE (relativa aos direitos comerciais sobre a propriedade intelectual). Considerar qual é o esquema motivador constitui procedimento imprescindível quando se trata da semântica de itens que integram um vocabulário especializado (científico: *id*, *superego*; *pulsões*; ou jurídico: *tomada de preços*, *consulta*, *pregão eletrônico*; ou esportivo: *pênalti*, *escanteio*, *grande área*). A novidade aqui está em sustentar que toda a significação linguística requer, para o seu processamento, o acesso a esquemas conceptuais independentes, cuja existência organiza as formas variáveis da convivência humana.

Essa asserção é menos óbvia quando se trata de lexemas evocadores de percepções ou de ações características das vivências humanas mais fundamentais. Nesses casos o senso-comum entre os semanticistas prefigura sua correspondência a “conceitos”, entendidos como entidades abstratas e desvinculadas da experiência na tradição das semânticas verificacionistas, versão hegemônica, mesmo quando implícita, dos tratamentos da significação. Afinal é verdade que “*a neve é branca*” se, e apenas se, a neve for branca. Sob esse ângulo, tanto o lexema *neve* como sua predicação, a palavra *branca*, têm como significação, respectivamente, e exclusivamente, suas condições de resolução referencial e suas condições de verificação no contexto (ou “estado-de-coisas”) relevante. É importante sublinhar que esse conjunto de condições semânticas (de referenciação ou de verificação) apresenta-se como

se fosse vazio de conteúdo experiencial (isto é, no caso, vazio de conteúdo sensorio-motor).

Não é o que sustenta a semântica cognitiva e, dentro dela, a semântica de frames, designada programaticamente como “*understanding semantics*”, isto é, semântica da compreensão (Fillmore, 1982, 1985). Nesta perspectiva, todo lexema evoca um frame, que pode, mesmo, corresponder ao esquema imagético exclusivamente evocável por aquele lexema: seja o caso do lexema *azul-claro*. O frame por ele evocado definirá valores específicos para os atributos do esquema MATIZ, LUMINOSIDADE e SATURAÇÃO, que são as dimensões em torno das quais se organiza geometricamente (como um sólido) o espaço da visibilidade cromática (conforme Palmer, 1999). Estamos longe, neste ponto, de um entendimento trivializante da categoria de frame, reduzido à sua dimensão antropológica. Queremos ressaltar que frames são categorias (sócio)cognitivas e que qualquer conceptualização, linguística ou não-linguística, processa-se como evocação (ou invocação, nos termos de Fillmore, 2008a) de um frame. É esta a posição assumida pela Teoria Neural da Linguagem (Feldman, 2006), segundo a qual todos os conceitos podem ser esquematizados como frames na forma de uma matriz de atributos (os Elementos do Frame) aos quais se associam situacionalmente valores parametrizados.

São, portanto, tratados como frames elementares os esquemas imagéticos, cinestésicos e executivos (os *X-schemas*) de que se ocupa a literatura cognitivista em linguística desde o trabalho pioneiro de Talmy (2000) até sua reivindicação teoricamente mais ambiciosa pela hipótese corporificada sobre a linguagem (em Regier, 1996; Grady, 1997; Naranayan, 1997; Lakoff e Johnson, 1999; Gibbs, 2006; Lakoff, 2008). Evidências neurobiológicas apontam para a realidade material da existência dos Elementos de um Frame como redes neurais especializadas (conforme Gallese e Lakoff, 2005): o estudo neurológico do gesto humano de PEGAR (GRASP) identifica os mesmos grupos neurais ativos tanto na realização motora dessa ação como na sua percepção (quando executada por outrem), delimitando claramente as funções de agente, paciente, modo (como se pega: com a mão aberta, com dois dedos, com os braços) e lugar (em que se situa o objeto “pegado”).

Linguisticamente, os Elementos de um Frame emergem na valência lexical e construcional das expressões evocadoras desse frame e correspondem, na designação mais tradicional da literatura, às Funções Temáticas da grade argumental dos núcleos lexicais – ou, numa perspectiva genealógica mais autêntica, aos Casos Semânticos, propostos pela Gramática de Casos (Fillmore, 1968). Mesmo aí cabe fazer uma distinção elucidativa. Os Elementos do Frame, na versão mais recente da teoria, são de fato Funções Microtemáticas, postuladas em relação ao frame a que se referem. Assim os lexemas *vender* e

*comprar*, que evocam ambos, com diferentes perspectivas, o frame do COMÉRCIO DE MERCADORIAS, têm como seus Elementos de Frame, respectivamente, as funções VENDEDOR e COMPRADOR: nisso diferem das abordagens mais tradicionais, que atribuiriam às mesmas funções o idêntico rótulo de agente.

A necessidade da postulação dessas Funções Microtemáticas em parte responde à inadequação das nomenclaturas anteriores, que, não tendo conseguido jamais apresentar um repertório consensual de Funções Temáticas, emprega algumas designações (Tema, por exemplo) com frustrante imprecisão. Ainda em termos estritamente descritivos, as listas de Casos Semânticos propostas jamais classificaram satisfatoriamente situações como as exemplificadas pelos constituintes negritos em sentenças como *Dunga* (como técnico) *substituiu o atacante Robinho no último jogo da seleção* versus *Nilmar* (como jogador) *substituiu o atacante Robinho no último jogo da seleção*.

De toda forma, a proeminência epistemológica da noção de frame como categoria cognitiva torna desejável que os Elementos dos Frames sejam identificados a partir das específicas situações conceptuais de que participam. Caberá às relações verificadas entre os frames (de Herança, de Pressuposição, de Perspectiva) representar as generalidades prevalentes entre os elementos dos esquemas conceptuais.

Um outro ponto importante é que os tratamentos anteriores das Funções Temáticas, na prática, restringiam-nas à definição das valências verbais. A semântica de frames elegera como tarefa tratar as relações combinatórias (sintático-semânticas) de todos os núcleos lexicais: Verbos, Nomes, Adjetivos, Preposições e Advérbios. Desse modo, a diferença entre *pegar um ovo* versus *pegar uma pedra* ou *pegar uma bola de futebol*, decorre não do frame de MANIPULAÇÃO (suscitado pelo Verbo *pegar*), mas sim dos frames respectivamente suscitados pelos Nomes *ovo*, *pedra* ou *bola de futebol*, que, a rigor, correspondem à estrutura de qualia associada a esses lexemas: é ao qual Constitutivo dos conceitos correspondentes que se seguirão as inferências de que se pega um ovo com delicadeza (para não quebrá-lo), mas se pega uma bola de futebol com firmeza (para que ela não escape).

Os comentários tecidos acima chamam nossa atenção para o fato de que, embora a semântica de frames surja como abordagem dedicada aos problemas da semântica lexical, é inegável sua importância também para as semânticas do discurso – basta levarmos em conta o elevado grau de sistematicidade que a postulação dos frames oferece aos processos conceptuais de inferenciação. Nesses termos, a semântica de frames converge complementarmente para as propostas da teoria do Léxico Gerativo (Pustejovsky, 1995) no seu intento de explicação da significação textual a partir das significações desencadeadas pelas unidades linguísticas que compõem o texto.

## O Projeto FrameNet

O Projeto FrameNet surge do cruzamento da semântica de frames com a lexicografia através da cooperação entre Fillmore e Atkins no princípio da década de noventa, e tem como resultado imediato a discussão já clássica do frame de RISCO e de suas realizações lexicais (Fillmore e Atkins, 1992, 1994). Dois componentes distintivos do Projeto emergem aí:

- (a) Primeiro, o propósito de organizar a descrição lexicográfica por frame: assim, ao frame de RISCO se associariam unidades lexicais como *risco*, *arriscar-se*, *arriscado*, *estar em risco*, *pôr em risco*, *correr o risco*, *perigo*, *cuidado*, *aventura*, *jogo*, *investimento*, *lucro*, *perigoso*, *perigosamente*, *aventuroso*, *conservador*, *cuidadoso*. A lista de exemplos contempla unidades lexicais cuja morfologia relaciona-se diretamente à designação linguística do frame (*risco*, *arriscar-se*) junto com outras inteiramente distintas (*jogo*, *lucro*, *perigo*, *cuidado*). Algumas dessas unidades são monolexêmicas; outras, multilexêmicas, incorporam verbos suporte como *pôr em risco*, *correr o risco*.
- (b) Outra característica importante é o uso de dados extraídos de corpus para descobrir todas as possibilidades combinatórias (ou valências) das unidades lexicais evocadoras de um frame – ou seja, todas as funções semânticas e propriedades gramaticais das expressões linguísticas que compõem as valências dessas unidades lexicais. Tal ângulo analítico permite revelar que o objeto direto do verbo *arriscar* pode instanciar três diferentes Elementos do Frame de RISCO, como ilustrado abaixo:

- (i) *Estou arriscando nossa amizade.* (BEM VALIOSO)
- (ii) *Estou arriscando levar uma descompostura.* (DANO)
- (iii) *Eu não arrisco telefonar para ela.* (AÇÃO)

A vantagem desse tipo de abordagem em relação à lexicografia tradicional é imediatamente verificável pela oferta de informações só disponíveis implicitamente nas definições dos dicionários tradicionais. Vide a descrição comparável do lexema *risco* disponibilizada pela versão eletrônica do dicionário Michaelis (Michaelis Dicionário Eletrônico, 2009):

*sm (ital rischio)* Possibilidade de perigo, incerto mas previsível, que ameaça de dano a pessoa ou a coisa. *R. bancário*, *Com*: o que decorre do negócio entre banqueiros ou entre o banco e os correntistas. *R. profissional*, *Dir*: perigo inerente ao exercício de certas profissões, o qual é compensado pela taxa adicional de periculosidade. *A risco de*, *com risco de*: em perigo de. *A todo o risco*: exposto a todos os perigos. *Correr risco*: estar exposto a.

Os Elementos do Frame de RISCO (POSSIBILIDADE DE DANO para o PROTAGONISTA de uma AÇÃO que pode gerar como consequência a PERDA DE UM BEM VALIOSO) estão, nesta definição, parcialmente presentes. Falta, porém, listagem mais abrangente das unidades lexicais que evocam a situação definida: há umas poucas expressões mencionadas (*a risco de*, *a todo o risco*, *correr risco*) e a sugestão de algumas *colocações* (*risco bancário*, *risco profissional*), sem exemplificação de seus usos; poderíamos, generosamente, inferir, da definição de *risco bancário*, que o lexema *especulação* é também associado ao lexema *risco*, mas esta inferência só a faremos porque sabemos Português e conhecemos intuitivamente o frame de RISCO: o dicionário não nos ajuda. Além disso, a ausência da enumeração das valências dessas expressões deixa de oferecer ao usuário as condições de distinguir, por exemplo, entre *risco de morte* e *risco de vida*.

Nos termos postos pelo próprio Fillmore (Fillmore, 2008a), o Projeto FrameNet visa superar limitações desse tipo, construindo um léxico

- (i) que seja baseado em evidência extraída de corpus;
- (ii) em que as unidades lexicais estejam associadas ao frame que evocam e que perfilam;
- (iii) que documente as propriedades combinatórias de cada uma dessas unidades lexicais por meio de sua anotação, definida pela metodologia do Projeto;

de modo que os verbetes desse léxico sumarizem, de forma sistemática, todas as informações mencionadas.

Na conclusão desse trabalho, estaremos nos aproximando do sonho do “dicionário ideal”, no qual cada um de nós, ao consultar uma palavra, seria remetido imediatamente para o frame que ela evoca, com todos os respectivos Elementos Componentes; veria, além disso, uma listagem de todas as valências desta palavra, suas possibilidades combinatórias sintáticas e semânticas, ilustradas por exemplos correspondentes. A consulta ainda nos ofereceria um conjunto de outras palavras que evocassem o mesmo frame, conectando-o com outros frames semanticamente relacionados.

É esse o horizonte que inspirou o Projeto FrameNet de Berkeley, com o objetivo “de criar um recurso lexical *on line* para o Inglês, baseado na semântica de frames e apoiado em evidência extraída de corpus”, conforme os termos com que apresentam o Projeto os próprios pesquisadores participantes (Ruppenhofer *et al.*, 2006). A página do Projeto na internet ([www.frame-net.icsi.berkeley.edu](http://www.frame-net.icsi.berkeley.edu)) noticia que sua base de dados, em desenvolvimento desde 1997, contém presentemente mais de dez mil unidades lexicais, das quais 6.100 anotadas de forma completa para mais de 825 frames, exemplificados por 135.000 sentenças. Os resultados, cuja liberação para



acesso público está na terceira versão, são hoje usados por centenas de pesquisadores em todo o mundo e servem de base para o desenvolvimento de recursos semelhantes para o espanhol (<http://gemini.uab.es:9080/SFNsite>), o alemão (<http://www.laits.utexas.edu/gframenet/>), o chinês e o japonês (<http://jfn.st.hc.keio.ac.jp/>), e, agora, também para o português. Uma outra aplicação vislumbrada do Projeto é subsidiar a rotulação semântica automática de texto corrido para diversos desenvolvimentos no Processamento da Linguagem Natural.

Há duas categorias teóricas usadas no Projeto que temos empregado informalmente, mas que exigem uma definição técnica. A primeira delas é a categoria de unidade lexical (UL), entendida como o pareamento de uma forma linguística (o significante) com a evocação de um único frame (o significado), nos termos de Cruse (1986). Isso requer que se distingam como diferentes ULs o verbo *pegar* em *pegar a xícara de café*, que evoca o frame de MANIPULAÇÃO, e o verbo *pegar* em *peguei a ideia que você me passou* ou *peguei a matéria da aula passada*, que evoca o frame de COMPREENSÃO. A forma linguística que participa da UL chama-se lema: lemas contrastam, como unidades analíticas, com palavras, que são variações morfológicas (inflexionais) do lema. Por exemplo, o lema *pegar* se instanciará nas palavras *peguei*, *pegar* (no infinitivo pessoal ou impessoal), *pegavam*, etc. Lemas também contrastam com lexemas, já que um lema pode consistir de um único lexema, como em *destampar* ou de mais de um lexema, como em *tirar a tampa*, ambos os quais evocam o frame de SEPARAÇÃO.

Os tratamentos cognitivistas da polissemia (por exemplo, Brugman, 1983; Lakoff, 1987, entre outros), que reivindicam o reconhecimento da relação metafórica entre as acepções de MANIPULAÇÃO e de COMPREENSÃO do lema *pegar* (vide Lakoff e Johnson, 1999, p. 240-241, sobre a metáfora conceptual COMPREENDER É PEGAR) não são contemplados neste estágio de desenvolvimento da FrameNet. Nada impede, entretanto, que as análises cognitivistas se prevaleçam das descrições oferecidas pela FrameNet para desenvolver e testar suas hipóteses: trabalhos representativos desta vertente teórica, como os reunidos sob a égide da Teoria Neural da Linguagem (Feldman, 2006; Lakoff, 2008) fazem-no usualmente.

A outra categoria que introduzimos sem definir é a da anotação. No Projeto FrameNet, há dois tipos de anotação: a anotação lexicográfica, que é a que praticamos na FrameNet Brasil, e a anotação de texto corrido. No caso da anotação lexicográfica, elegemos um frame, ou uma família de frames relacionados, identificamos ULs que lhe(s) sejam associadas, e pesquisamos em corpus exemplos que atestem a variedade de condições combinatórias destas ULs. A anotação de texto corrido

procede pela anotação de todas as sentenças que constituem o texto escolhido. Nas palavras de Ruppenhofer *et al.* (2006, p. 7), os dois tipos de anotação se diferenciam porque, no caso da anotação de texto corrido, “as sentenças a serem anotadas são escolhidas, por assim dizer, pelo autor do texto”. Nos dois casos, as anotações distinguem a UL alvo, evocadora do frame, e os constituintes que participam da valência desse uso, os quais têm suas características registradas graficamente na forma de camadas que especificam, no mínimo, sua função semântica, ou seja, o Elemento do Frame (EF) a que correspondem; e sua descrição sintática, em termos de sua classe sintática, o Tipo de Sintagma (TS) e de sua Função Gramatical (FG). Exemplificamos abaixo:

	Os alunos	recortaram	do jornal	todas as notícias sobre a epidemia.
EF	Agente		Item	Pedaços
GF	Ext		Dep	Obj
TS	SN		SP	SN

A instância acima anota como UL alvo a palavra *recortaram*, que evoca o frame de CORTE: os constituintes que ocorrem com a UL vão marcados em cores<sup>2</sup> cuja identidade tem função coindexadora. Vemos aí que o Sintagma Nominal *os alunos*, que desempenha a Função Gramatical (Argumento) Externo (categoria funcional mais heterogênea que a de Sujeito, conforme Ruppenhofer *et al.*, 2006, p. 92), é rotulado com a função semântica de Agente. E assim por diante.

As categorias da anotação procedem em parte da descrição dos frames: é o caso dos EFs, específicos para cada frame. Já as categorias sintáticas (GFs e TSs) procedem da análise gramatical assumida pelo Projeto, que, declarando-se programaticamente “suprateórico”, assume de forma reconhecível os pressupostos das teorias sintáticas construcionistas e não-derivacionais, tais como a Teoria da Sintaxe Mais Simples (Culicover e Jackendoff, 2005), ou da Head-Driven Phrase Structure Grammar, HPSG (Sag *et al.*, 2003) e, principalmente, da própria Gramática das Construções, tal como vem sendo desenvolvida por Fillmore, Kay e colaboradores, ao longo dos últimos vinte anos. O acervo das categorias analíticas nesse caso é suprido pelos capítulos 4, 5 e Apêndice de Ruppenhofer *et al.* (2006), súmula teórica e manual de procedimentos metodológicos do Projeto, designado jocosamente como “the book”, disponível para ser baixado como arquivo eletrônico da própria página da FrameNet.

A anotação, mesmo de dados do inglês, língua que é o objeto originário do Projeto, não procede sem que alguns percalços sejam enfrentados. (A anotação em

<sup>2</sup> Nota das editoras: as cores foram substituídas por graus de sombreamento em cinza, devido a limitações gráficas relacionadas com a impressão da revista.

português lida com problemas específicos que abordaremos na próxima seção). A falta de isomorfismo entre as estruturas gramaticais e as estruturas semânticas das línguas humanas requer que sejam previstas, por exemplo, soluções para situações em que se omite sintaticamente um Elemento do Frame (EF).

Um EF pode ser omitido quando faz referência a situações previamente mencionadas no discurso: trata-se, nesse caso, de uma anáfora nula, designada no Projeto como Instanciação Nula Definida (IND). Considere-se, por exemplo, a sentença *Ele saiu e **comprou** umas cervejas* [IND<sub>comprador</sub>], em que, para a UL alvo *comprou*, o EF *comprador* ocorre como IND. Um EF pode também ter sua omissão lexicalmente licenciada quando o uso considerado da UL alvo admite uma interpretação genérica, ou, algumas vezes catafórica, do EF em questão. Seja o caso da sentença *Eu passei a tarde **costurando***: a UL *costurando* tem dois EFs omitidos; o Agente é um caso de IND (mencionado na oração principal), mas o objeto do ato de costurar é um caso de INI, isto é, uma Instanciação Nula Indefinida, já que parece pragmaticamente irrelevante na situação especificar que peças tenham sido costuradas. Uma terceira alternativa de omissibilidade de um EF é dada construcionalmente: é o caso da omissibilidade do agente nas construções passivas. Tais EFs, designados como Instanciações Nulas Construcionais (INC), são anotados nesta condição.

Outra manifestação de falta de isomorfismo entre estruturas formais e estruturas semânticas, que requer solução no Projeto, é o fenômeno designado *mismatch* (ou desencontro) na literatura construcionista e nos estudos de processos de gramaticalização (Francis e Michaelis, 2003; Michaelis, 2004, 2005; Traugott, 2007). São casos notórios de desencontro as divergências entre as “regências” sintática e semântica de um sintagma: a hipótese *default* do isomorfismo é que o “regente” (ou núcleo) sintático de um sintagma seja também o “regente” semântico da expressão. Ocorrem, entretanto, desencontros com núcleos verbais e, também, com núcleos nominais.

Exemplificam esta situação as cópulas (que, na tradição gramatical, são tratadas como introduzindo um “predicado nominal”, à falta de melhor designação para o desencontro). Seja a sentença *Ela parece doente*. A cópula *parece*, no Projeto FrameNet, jamais será tratada como UL alvo; ao invés, será anotada como cópula na camada Adj, uma vez que a UL evocadora do frame que organiza semanticamente esta sentença é o adjetivo *doente*.

Além disso, merecem destaque os verbos suporte (Mel’cuk e Wanner, 1996; Moura Neves, 2000), um dependente dos quais introduz a principal informação semântica da sentença. Segundo Fillmore (2008b), os verbos suporte classificam-se como intransitivos (*O vento soprava com violência*) ou transitivos (*ela deu um grito, soltou uma risada, deu uma declaração, fez um juramento, tomou a decisão, teve uma briga com o irmão*), e, ainda,

transitivos diretos e indiretos (*me deu a palavra, te fez uma promessa*). Os verbos suporte podem introduzir, nas construções em que ocorrem, informação causativa (*aula me põe com dor de cabeça* versus *estou com dor de cabeça*), informação aspectual (*ele começou / acabou a aula as dez em ponto*), informação sobre o ponto de vista prevalente (*dar uma ajuda* versus *receber uma ajuda*), ou informação sobre o registro comunicativo (*dar queixa* versus *registrar queixa*). O Projeto FrameNet anota os verbos suporte, tal como a cópula, na camada designada pela classe sintática da UL portadora-do-frame, que é frequentemente um nome. (Haveria muito mais a dizer sobre expressões suporte e sobre desencontros sintático-semânticos, mas não é esse o objeto deste artigo.)

Também há desencontros envolvendo a regência dos nomes em expressões do tipo N<sub>1</sub> de N<sub>2</sub>, caso em que o nome subordinado N<sub>2</sub> é que mantém a relação semântica com o frame instanciado pela UL-alvo. Por exemplo, na sentença *Bebi uma taça de vinho*, o frame evocado pela UL *bebi* toma como EF a substância instanciada pelo N<sub>2</sub> *vinho* e não pelo N<sub>1</sub> *taça*. O Projeto designa esse tipo especial de nomes, cujo dependente é selecionado pelo verbo, como nomes transparentes (*penca de bananas, dúzia de ovos, lata de cerveja, monte de gente, este tipo de problema, porção de batatas fritas, gomo de laranja*). Os nomes transparentes funcionam como classificadores nominais (de quantidade, tipo, configuração...) e, também sobre eles, haveria muito mais a dizer.

Outra situação de desencontro tratada na FrameNet tem a ver com a incorporação morfológica de um EF na própria UL portadora do frame e alvo da anotação: por exemplo, a UL *fatiar*, evocadora do frame de CORTE, usada na sentença *Ela fatiou o salame*, terá o EF<sub>PEDAÇOS</sub> (*fatias*) anotado como estando incorporado ao próprio núcleo lexical. Esta situação de omissibilidade de um EF não se confunde com aquelas (IND, INI e INC), que acima mencionamos, porque, no caso das incorporações, o EF incorporado pode instanciar-se sintaticamente (e.g. *Ele fatiou* [INC: PEDAÇOS] *o salame* [em pedaços irregulares<sub>PEDAÇOS</sub>]).

Um tipo adicional de complexidade refere-se à anotação de constituintes que mantêm com o frame dupla vinculação semântica, isto é, instanciam, ao mesmo tempo, dois elementos de um dado frame. Por exemplo, na sentença *Eu admiro a sua atitude*, o constituinte *a sua atitude* apresenta a confluência sintática de dois elementos do frame AVALIAÇÃO: o AVALIADO e a RAZÃO da AVALIAÇÃO. Considere-se que uma paráfrase aproximada dessa sentença poderia distinguir sintaticamente os dois EFs: *Eu admiro* [Você AVALIADO] [por esta atitude<sub>RAZÃO</sub>]. Para anotar esse tipo de situação, acrescenta-se uma segunda camada de elementos do frame à camada EF, prevista regularmente.

A anotação lexicográfica, tal como sumariamente a apresentamos, gera uma base de dados acessível na página do Projeto, que tem, para o usuário, a seguinte configuração:

- (i) há uma lista de frames, descritos individualmente por sua definição e pelo conjunto pertinente de elementos do frame, tanto nucleares como periféricos: cada um dos EFs é, por sua vez, definido e ilustrado com exemplificações anotadas, extraídas de corpus;
- (ii) há uma lista de Unidades Lexicais associadas a cada frame, mencionadas com sua classe sintática (e.g. *slice,n.*; *slice,v.*), as quais, por sua vez, remetem aos contextos em que ocorreram “de forma mais saliente”: clicado um desses contextos de ocorrência, emergirá para nós o conjunto de exemplificações anotadas que o ilustram;
- (iii) os exemplos aparecem anotados em camadas conforme é previsto no Projeto: identificada a UL alvo, são classificados os constituintes que, em relação a ela, mantêm relações semânticas (como EFs) e sintáticas (Tipo de Sintagmas e Função Gramatical); é possível, porém, prescindir da visualização da anotação em camadas e solicitar a exibição das exemplificações por tipo de rotulação (apenas os EFs, ou os TSs, ou as FGs);
- (iv) os verbetes correspondentes a cada UL contêm a identificação do frame que evocam, mais uma tabela contendo a discriminação dos Elementos do Frame que com ela co-ocorrem e suas respectivas realizações sintáticas: as casas dessa tabela são vinculadas aos exemplos anotados que as ilustram;
- (v) a outra parte constituinte do verbete é uma nova tabela que discrimina todos os padrões de valência levantados para a UL considerada, sendo cada padrão conectado aos exemplos anotados que o ilustram: no caso da anotação de alvos nominais, acrescenta-se a listagem das ULs regentes ou suportes que com eles co-ocorreram.

Como se observa, trata-se de uma base de dados bastante complexa, que contém todos os cruzamentos entre frames e ULs considerados relevantes pelo Projeto.

Por último, é necessário elucidar observação que fizemos anteriormente sobre as vantagens de tratar como funções microtemáticas as relações semânticas que correspondem aos elementos de um frame. Na ocasião mencionamos que as generalizações supostamente perdidas ao rotular determinados elementos de um frame como comprador ou como vendedor, ao invés de usar o rótulo

mais geral agente, seriam compensadas com a postulação das relações interframes. De fato, os frames COMÉRCIO – COMPRAR e COMÉRCIO – VENDER, herdaram EFs do frame TRANSFERÊNCIA, que, através do frame de TROCA, herda elementos, por sua vez, do frame AGIR – INTENCIONALMENTE, o qual inclui agente como um de seus EFs. Assim, a generalização, que é complexa, fica preservada e, mais que isso, a generalização fica motivada de forma sistemática na rede semântica que é a FrameNet (em português, “rede de frames”, ou “rede de esquemas conceptuais”).

A Figura 1, extraída da base de dados da FrameNet, organizada pela ferramenta FrameGrapher, nos dá uma ideia desse tipo de sistematização.

As linhas em vermelho no gráfico original<sup>3</sup> (que ligam “Transfer” a “Commerce\_goods-transfer” e “Commerce\_money-transfer”) representam as relações de herança; as linhas azuis (que ligam “Commercial\_transactions” a “Commerce\_goods-transfer” e “Commerce\_money-transfer”) representam as relações de constituição de um frame em subframes; as linhas verdes (que ligam “Commerce\_goods-transfer” a “Commerce\_buy” e “Commerce\_sell” e ligam “Commerce\_money-transfer” a “Commerce\_collect” e “Commerce\_pay”), no caso, indicam diferenças de perspectiva sobre uma mesma cena conceptual. Muito mais, aqui também, poderíamos acrescentar, mas não é esse o objetivo do presente artigo.

## O projeto FrameNet Brasil

### Identificação institucional

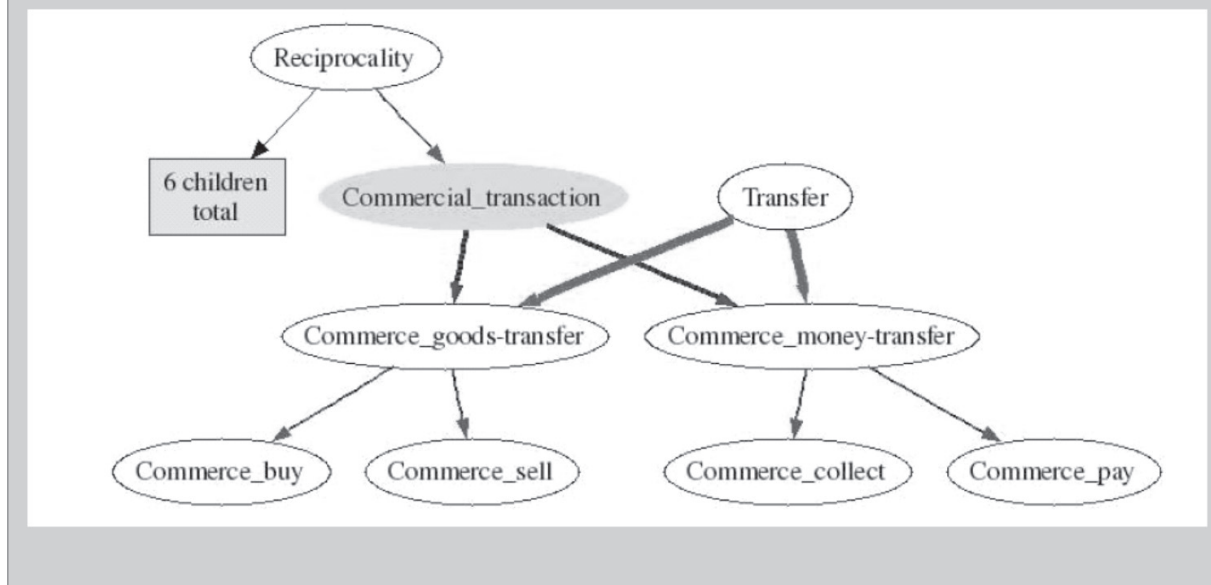
O Projeto FrameNet Brasil originou-se de contato estabelecido por pesquisadores brasileiros da UFJF com a equipe do Projeto FrameNet na Califórnia, no ano de 2007, ocasião em que a autora deste texto, sob patrocínio da CAPES realizava estágio como Visiting Scholar ao Departamento de Linguística da Universidade da Califórnia, em Berkeley. No segundo semestre daquele ano, o Professor Carlos Subirats, principal investigador do Projeto Spanish FrameNet, visitou a Universidade Federal de Juiz de Fora, quando liderou oficina sobre sua experiência de implantar, aplicado ao espanhol, um projeto lexicográfico análogo àquele que vinha sendo desenvolvido para o inglês.

Tivemos também como participante desta oficina a professora Rove Chisman, da UNISINOS, que, nessa instituição, coordena a implantação do Projeto Frame-Corp, uma proposta de aplicação da semântica de frames à anotação de corpus, e que, com esse propósito, também utiliza a base de dados do Projeto FrameNet de Berkeley.

<sup>3</sup> Nota das editoras: por motivo alheio a nossa vontade, não foi possível adaptar o gráfico original a uma versão em preto e branco. Por isso, mantivemos a menção às cores do original, mas acrescentamos, entre parênteses, a descrição das ligações das respectivas linhas, para que o gráfico possa ser lido também nesta versão em preto e branco.



# Commercial Transactions



**Figura 1:** Sistematização feita pelo FrameGrapher para Transações Comerciais.

**Figure 1:** FrameGrapher sistematization for Commercial Transactions.

Sob esse ângulo, os projetos FrameCorp e FrameNet Brasil são complementares, na medida em que partem da mesma base de dados, para proceder, o primeiro, a anotações de texto corrido, enquanto o segundo pratica a anotação lexicográfica. Ambos os projetos têm procurado manter contatos acadêmicos entre si.

O Projeto FrameNet Brasil, que conta, desde janeiro de 2009, com financiamento da FAPEMIG, é campo que subsidia presentemente a pesquisa para elaboração de dissertações de mestrado e teses de doutorado, além de congregar o trabalho de bolsistas de iniciação científica e de apoio profissional. Entre as metas do Projeto, está a disponibilização do acesso público à página bilíngue (em Português e em Inglês).

## Corpus do Projeto FrameNet Brasil

A base de dados do corpus do Projeto FrameNet Brasil, em construção, é presentemente constituída pela combinação dos seguintes corpora, todos caracterizados por permitir acesso público, e que representam usos do Português Europeu e do Português do Brasil:

- (i) *Subcorpus do Corpus de Referência do Português Contemporâneo (CRPC) de linguagem oral;*
- (ii) *Português Falado – Variedades Geográficas e Sociais:* corpus constituído por 86 textos que vão

de conversas informais até programas radiofônicos de diversos países que falam português;

- (iii) *Museu da Pessoa:* corpus de 107 entrevistas transcritas pelo Núcleo Português do Museu da Pessoa no âmbito dos seus projetos, mais seis entrevistas transcritas pelo Museu da Pessoa brasileiro;
- (iv) *NILC/São Carlos:* contém textos brasileiros do registro jornalístico (do qual se originou o CETENFolha), didático, epistolar e redações de alunos;
- (v) *CETEMPúblico:* contém cerca de 190 milhões de palavras extraídas do diário *Público* no período de 1991 a 1998;
- (vi) *CHAVE:* contém textos jornalísticos dos jornais *Público* e *Folha de São Paulo* dos anos de 1994 e 1995;
- (vii) *CONDIVport:* contém textos dos domínios do futebol, do vestuário/moda e da saúde das décadas de 50, 70 e 2000 de jornais e revistas portuguesas e brasileiros;
- (viii) *DiaClav:* contém mais de 12 mil artigos de edições *on line* de diversos jornais diários regionais do centro de Portugal, nomeadamente o *Diário de Coimbra*, *Diário de Leiria*, *Diário de Aveiro* e *Viseu Diário*;

- (ix) *Natura/Minho*: corpus jornalístico contendo texto do jornal regional português *Diário do Minho*, criado no âmbito do projeto *Natura*;
- (x) *Avante*: contém textos do semanário “*Avante!*”, órgão oficial do Partido Comunista Português;
- (xi) *ANCIB*: corpus criado a partir de mensagens enviadas para a lista homônima da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (até Novembro de 2003) e para a lista *abarretol*, após essa data;
- (xii) *CoNE*: corpus constituído de mensagens não-endereçadas em português;
- (xiii) *Clássicos LP/Porto Editora*: contém textos literários portugueses cedidos pela Porto Editora;
- (xiv) *Vercial*: corpus de obras literárias de autores portugueses do século XIX e início do século XX;
- (xv) *ENPCPUB*: corpus constituído de traduções para português de textos literários em inglês;
- (xvi) *ECI-EBR*: corpus criado pela ECI (European Corpus Initiative), baseado no *Borba-Ramsey*. É uma seleção de excertos de obras brasileiras, contendo pelo menos discurso literário, didático e oral cuidado (discursos políticos).

Os dois primeiros corpora foram disponibilizados pelo CRPC da Universidade de Lisboa e todos os demais estão acessíveis no site *Linguatca* e estão sendo usados com conhecimento e permissão dos pesquisadores responsáveis.

Os corpora acima totalizam pouco mais de 280 milhões de palavras, constituindo, no momento, uma base menor do que o corpus da *FrameNet* para o Espanhol, que reúne cerca de 350 milhões de itens. Nosso Projeto está no momento empenhado em acrescentar mais 30 milhões de itens em dados de língua oral.

A Tabela 1 mostra a distribuição de itens por gênero textual nos corpora enumerados. As classificações por gênero acompanham as descrições feitas pelos próprios corpora, nas suas autodefinições, razão pela qual a tabela inclui a problemática distinção entre os “gêneros” *didático-científico* e *universitário*. O Gráfico 1 correspondente à Tabela 1 apresenta a configuração.

É evidente o desequilíbrio desse corpus, especialmente no que concerne à modalidade oral do uso linguístico. Há outros desequilíbrios, porém: mesmo no que concerne à modalidade escrita, o predomínio do gênero jornalístico é avassalador. Estamos, no presente estágio do Projeto, trabalhando para incorporar mais dados de língua falada e de escrita oralizada (nos termos de Marcuschi, 2000, p. 15-43). A dificuldade diagnosticada alcança, pelo que pudemos constatar, outras versões do Projeto *Frame-*

**Tabela 1:** Distribuição de itens por gênero textual.

**Table 1:** Item distribution per textual genre.

Gênero	Total
Oral	812.446
Didático	1.388.660
Jurídico	761.852
Literário	12.179.584
Técnico e Científico	1.767.565
Jornalístico	262.122.234
Universitário	1.027.908
Mensagem eletrônica	1.887.528

*Net*. Necessidades adicionais de refinamento concernem à distribuição dos dados quanto à distinção entre as versões europeia e brasileira do português e à sua identificação em termos diacrônicos.

Nossa pretensão é de que o corpus da *FrameNet* Brasil, com os enriquecimentos que, neste momento, estamos negociando, fique, à disposição da comunidade dos pesquisadores em língua portuguesa à semelhança de outros projetos de análoga natureza (entre eles, destacadamente, o *Linguatca*).

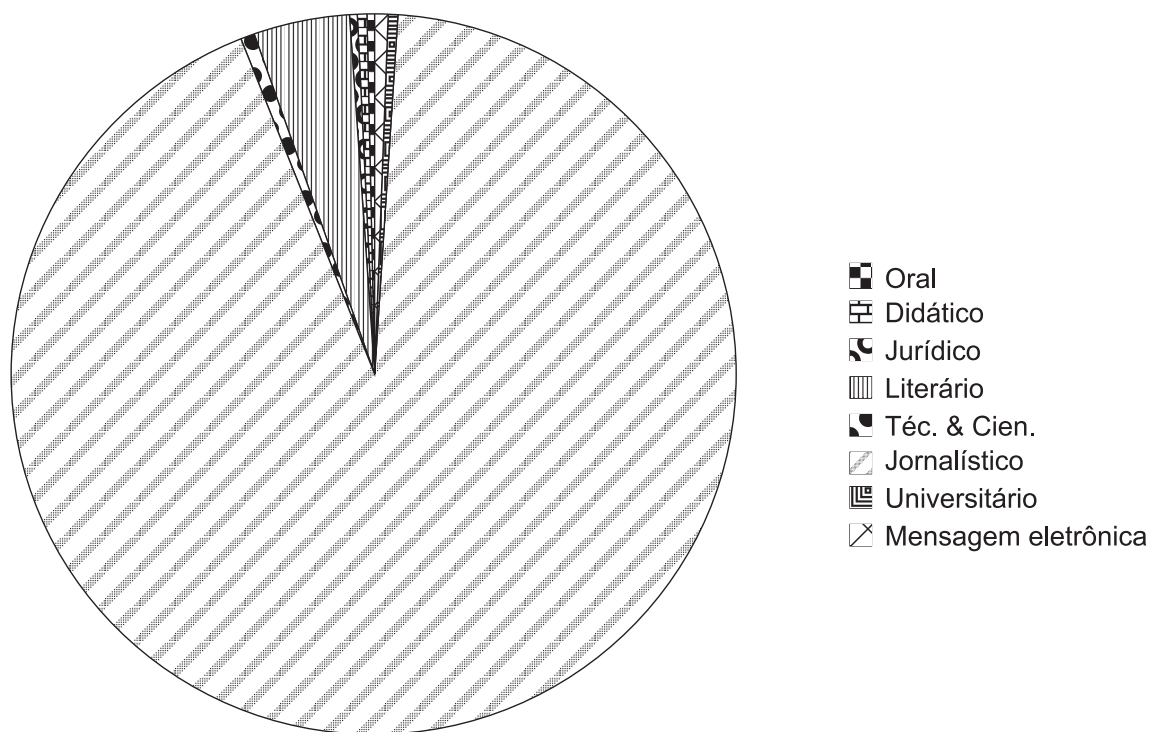
### O foco da presente etapa de desenvolvimento

Poderíamos ter iniciado nossa pesquisa por um frame de *CONEXÃO* (“*the case of attaching*”), acompanhando Fillmore *et al.* (2003) na sua detalhada descrição da metodologia de implementação do Projeto *FrameNet*. Inspiramo-nos, ao invés, numa investigação em curso no Instituto Max Plank sobre os padrões de lexicalização dos esquemas conceituais de “desconstituição de uma unidade física”, verificados numa amostra formada por dezessete línguas geneticamente dissimilares (conforme Majid *et al.*, 2007).

Tais padrões de lexicalização são ilustrados por lexemas do tipo *cut/break*, em inglês, sobre os quais existe copiosa literatura: não obstante a proximidade semântica que se possa atribuir a esses lexemas, é notável a diferença em suas instâncias sintáticas – enquanto *break* pode ocorrer como verbo transitivo (*He broke the stick*) ou intransitivo (*The stick broke*), *cut* só admite sua instânciação como verbo transitivo (*He cut the cloth*. \**He cut*. \**The cloth cut*.). Do nosso ponto de vista, tal diferença sintática é homóloga à evocação de frames causativos (pelas construções transitivas) ou incoativos (pelas correspondentes construções intransitivas).

Há uma quantidade de ULs que evocam esses cenários de “desintegração da unidade física”: desde logo, distinguimos entre cenários de ação intencional em que se controla o resultado da ação (por exemplo, cortar a cebola em rodelas) e cenários em que o resultado produzido corresponde a pedaços de formato e constituição aleatórios (por exemplo, espatifar a louça

## FrameNet Brasil - Corpus



**Figura 2:** Gráfico da distribuição de itens por gênero textual.

**Figure 2:** Graph of item distribution per textual genre.

no chão). A primeira classe desses cenários corresponde ao frame de CORTE, que é necessariamente agentivo. A segunda classe de situações corresponde à família dos frames de FRAGMENTAÇÃO, que terão uma variante agentiva (causativa) e uma variante incoativa. Em ambos os casos, a “desintegração” se dá com perda material da massa física do objeto afetado. Além dessas duas classes de eventos, identificamos uma terceira, distinta das precedentes pela sua reversibilidade (por exemplo, tirar a tampa do açucareiro): nesse caso, não há perda material do item, e a situação que se verifica é aquela em que uma parte é temporariamente afastada do todo. Trata-se, aqui, do frame de SEPARAÇÃO.

Esses frames, cuja lexicalização desejamos estudar, são objeto das dissertações de Mestrado das Professoras Renata Cristina de Barros Vieira Marques (CORTE), Paula Zagotta de Oliveira (FRAGMENTAÇÃO) e Gabriela da Silva Pires (SEPARAÇÃO). O trabalho em curso tem sido o de identificar ULs (verbos, nomes e adjetivos) evocadoras desses frames e anotá-las como alvo nos termos propostos para o Projeto. Dezembro de 2009 é a data da conclusão desta etapa da implantação da FrameNet Brasil, com a produção do léxico correspondente, além da descrição dos frames especializados, associados às Unidades Lexicais estudadas (por exemplo, as ULs *serrar*,

*fatiar e recortar* evocam todas o frame de CORTE, mas, além disso, evocam cenas conceituais específicas que as distinguem como formas variantes de CORTAR.)

### O processo de anotação lexicográfica

O processo de anotação se inicia pela preparação dos subcorpora relevantes. Após a produção de concordâncias para as ULs alvos (empregando softwares de busca lexical como o SketchEngine ([www.sketchengine.co.uk/](http://www.sketchengine.co.uk/)) ou a ferramenta disponibilizada pelo projeto AC/DC ([www.linguateca.pt/ACDC/](http://www.linguateca.pt/ACDC/))), passamos a um processo de classificação (“limpeza”) dos dados, finalizado o qual obtemos uma base de usos da UL na acepção (de “perda da unidade física”) que estamos estudando. Na prática, isso implica descartar todos os usos figurativos das ULs, os quais são, não surpreendentemente, muito numerosos: a abundância desses usos é consistente com hipótese corrente na Linguística Cognitiva, segundo a qual as metáforas primárias empregam domínios fonte originários de experiências físicas, tais como *cortar*, *quebrar* ou *separar alguma coisa*.

Quando os subcorpora ultrapassam uma dimensão manejável para a anotação não-automática (que estipulamos em quatrocentas instâncias), definimos, para anota-

ção, uma amostra desse subcorpus, empregando para isso procedimentos estatísticos convencionais e trabalhando com um grau de confiança de 95%. Apesar de recorrermos à estatística para a preparação dos subcorpora, nossos verbetes não incluem informação sobre frequência dos usos estudados, acompanhando também nesse ponto o Projeto FrameNet de Berkeley.

O estudo da maior parte dos subcorpora até agora não vem requerendo recurso à amostragem: a natureza de nossa base de dados, com domínio tão flagrante do discurso jornalístico, acaba gerando dificuldades para que se apure quantidade suficiente de usos que nos interessam: de fato, não é fácil, com esse corpus, levantar uma base quantitativamente relevante de usos de uma UL como *ralar* no sentido de CORTE. Circunstâncias como estas é que têm contribuído para nossa percepção de que é imprescindível a ampliação do corpus através da incorporação de dados de língua oral.

Um esforço significativo que temos feito é o de revisão e adaptação da rotulação gramatical dos constituintes que correspondem aos elementos do frame evocado pela UL alvo. As funções gramaticais, pela generalidade que caracteriza esse tipo de categoria sintática, não são problemáticas, mas os Tipos de Sintagmas (TSs) divergem expressivamente entre o português e o inglês: basta lembrar a relevância das construções gerundiais (oracionais ou sintagmáticas), muito mais diversificadas e frequentes em inglês do que em português; por outro lado, em português, as construções infinitivas (pessoais e impessoais) são muito mais importantes do que na gramática do inglês. Respeitando as premissas metodológicas do Projeto, de aderir a uma descrição “simples” e não-derivacional, estamos, no momento, experimentando um repertório de categorias, que serão divulgadas na página do Projeto.

Uma outra especificidade de nossos dados, que ainda merecerá uma checagem mais rigorosa, é a frequência, aparentemente alta, de Instanciações Nulas Definidas (INDs) de EFs em condições catafóricas: Fillmore (2008b) menciona, na sua análise dos dados do inglês, uma tendência de interpretação catafórica para as Instanciações Nulas Indefinidas (INIs) dos EFs. Indicamos esse fato como um fenômeno, entre numerosos outros, que esse tipo de pesquisa, necessariamente comparativa, ensejará abordar.

Por último, um ponto que tem chamado a nossa atenção é a necessidade de rever as relações interframes, apresentadas no Projeto FrameNet do inglês através da ferramenta FrameGrapher. Embora seja intuitivamente esperável a postulação de relações conceptuais, mesmo que indiretas, entre os frames de CORTE, FRAGMENTAÇÃO e SEPARAÇÃO (intuição motivadora do Projeto do Instituto Max Plank), não há nenhuma conexão proposta entre esses frames na versão original da FrameNet. Tal característica é, na verdade, recorrente: as relações postuladas ora o são num nível extremamente genérico

(referentes, por exemplo, a tipos de eventos), ou, quando mais detalhadas, restringem-se aos esquemas conceptuais, cujo estudo inspirou a semântica de frames (por exemplo, a cena comercial ou os verbos de julgamento). Muito há, pois, a fazer para que tenhamos, para as várias versões do Projeto FrameNet, uma ontologia abrangente e empiricamente fundada na semântica linguística.

Mesmo no estágio inicial em que se encontra nosso projeto, em processo de construção de seu *corpus* e de depuração de suas categorias analíticas, podemos vislumbrar-lhe, a curto prazo, um universo de fecundas descobertas científicas, com potencial para apoiar desenvolvimentos tanto na área de lexicografia como das tecnologias da informação.

## Referências

- BRUGMAN, C. 1981. *Story of over*. Berkeley, CA. Dissertação de mestrado. University of California, 127 p.
- CRUSE, D.A. 1986. *Lexical semantics*. Cambridge, Cambridge University Press, 310 p.
- CULLICOVER, P.W.; JACKENDOFF, R. 2005. *Simpler syntax*. New York, Oxford University Press, 589 p.
- FELDMAN, J. 2006. *From molecule to metaphor: A neural theory of language*. Cambridge, MIT Press, 357 p.
- FILLMORE, C.J. 1968. The case for case. In: E. BACH; R. HARMS (eds.), *Universals in linguistic theory*. New York, Holt, Rinehart and Winston, p. 1-88.
- FILLMORE, C.J. 1975. An alternative to checklist theories of meaning. In: C. COGEN; H. THOMPSON; G. THURGOOD; K. WHISTLER (eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley, p. 23-131.
- FILLMORE, C.J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, 280:20-32.
- FILLMORE, C.J. 1977. The case for case reopened. In: P. COLE; J. SADDOCK (eds.), *Grammatical relations*. New York, Academic Press, p. 59-81.
- FILLMORE, C.J. 1982. Frame semantics. In: THE LINGUISTICS SOCIETY OF KOREA, *Linguistics in the morning calm*. Seoul, Hashin, p. 111-137.
- FILLMORE, C.J. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 1(6):222-254.
- FILLMORE, C.J. 2008a. FrameNet: The lexicon. Disponível em: <http://www.hf.uib.no/forskerskole/CxG.html>. Acesso em: 29/11/2008.
- FILLMORE, C.J. 2008b. Beyond Simple FrameNet: Some subtleties. Disponível em: <http://www.hf.uib.no/forskerskole/CxG.html>. Acesso em 29/11/2008.
- FILLMORE, C.J.; ATKINS, B.T.S. 1992. Towards a frame-based lexicon: The case of RISK. In: A. LEHRER; E. KITTAY (eds.), *Frames and fields*. Hillsdale, Lawrence Erlbaum, p. 75-102.
- FILLMORE, C.J.; ATKINS, B.T.S. 1994. Starting where the dictionaries stop: The challenge for computational lexicography. In: B.T.S. ATKINS; A. ZAMPOLLI (eds.), *Computational approaches to the lexicon*. Oxford, Oxford University Press, p. 349-393.
- FILLMORE, C.J.; PETRUCK, M.R.L.; RUPPENHOFER, J.; WRIGHT, A. 2003. FrameNet in action: the case of Attaching. *International Journal of Lexicography*, 16(1):297-332.
- FRANCIS, E.; MICHAELIS, L.A. (eds.). 2003. *Mismatch: Form-function incongruity and the architecture of grammar*. Stanford, CSLI Publications, 367 p.
- GALLESE, V.; LAKOFF, G. 2005. The brain's concepts: the role of sensory-motor systems in conceptual knowledge. *Brain and Language*, 89:385-392.



- GIBBS, R. 2006. *Embodiment and cognitive science*. Cambridge, Cambridge University Press, 337 p.
- GOFFMANN, E. 1974. *Frame analysis: Essays on the organization of experience*. New York, Harper, 289 p.
- GRADY, J. 1997. *Foundations of meaning: Primary metaphors and primary scenes*. Berkeley, CA. Dissertação de Ph.D. University of California, 306 p.
- LAKOFF, G. 1977. Linguistic gestalts. In: REGIONAL MEETING OF THE CHICAGO LINGUISTICS SOCIETY, 13, Chicago, 1977. *Anais...* Chicago Linguistics Society, p. 236-287.
- LAKOFF, G. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, University of Chicago Press, 614 p.
- LAKOFF, G. 2008. The neural theory of metaphor. In: R. GIBBS (ed.), *The Cambridge handbook of metaphor and thought*. Oxford, Oxford University Press, p. 17-38.
- LAKOFF, G.; JOHNSON, M. 1999. *Philosophy in the flesh*. New York, Basic Books, 624 p.
- LANGACKER, R. 1982. Space grammar, analyzability, and the English passive. *Language*, 58:22-80.
- LANGACKER, R. 1987. *Foundations of cognitive grammar. Theoretical prerequisites. Vol. 1*. Stanford, Stanford University Press, 516 p.
- LANGACKER, R.; MUNRO, P. 1975. Passives and their meaning. *Language*, 51:789-830.
- MAJID, A.; BOWERMAN, M.; VAN STADEN, M.; BOSTER, J.S. 2007. The semantic categories of cutting and breaking events: A crosslinguistic perspective. *Cognitive Linguistics*, 18(2):133-152.
- MARCUSCHI, L.A. 2000. *Da fala para a escrita*. São Paulo, Cortez, 133 p.
- MEL'CUK, I.; WANNER, L. 1996. Lexical functions and lexical inheritance for emotion lexemes in German. In: L. WANNER (ed.), *Lexical functions in lexicography and natural language processing*. Amsterdam/Philadelphia, Benjamins, p. 209-278.
- MICHAELIS DICIONÁRIO ELETRÔNICO. 2009. Disponível em <http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues>. Acesso em 03/07/2009.
- MICHAELIS, L. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics*, 15:1-67.
- MICHAELIS, L. 2005. Entity and event coercion in a symbolic theory of syntax. In: J.-A. OSTMAN; M. FRIED (eds.), *Construction grammar(s): Cognitive grounding and theoretical extensions. constructional approaches to language, Volume 3*. Amsterdam, Benjamins, p. 45-87.
- MOURA NEVES, M.H. 2000. *Gramática de usos do português*. São Paulo, Editora UNESP, 1037 p.
- NARANAYAN, S. 1997. *embodiment in language understanding: Sensory-motor representations for metaphoric reasoning about event descriptions*. Berkeley, CA. Dissertação de Ph.D. University of California, 267 p.
- PALMER, S. 1999. *Vision science: Photons to phenomenology*. Cambridge, MIT Press, 810 p.
- PUSTEJOVSKY, J. 1995. *The Generative Lexicon*. Cambridge, MIT Press, 298 p.
- REGIER, T. 1996. *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MIT Press, 220 p.
- RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M.; JOHNSON, C.; SCHEFFCZYK. *FrameNet II: Extended theory and practice*. Disponível em: <http://framenet.icsi.berkeley.edu/>. Acesso em: 03/07/2009.
- SAG, I.; WASOW, T.; BENDER, E. 2003. *Syntactic theory: A formal introduction*. 2ª ed., Stanford, CSLI. 608 p.
- SCHANK, R.; ABELSON, R. 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, Lawrence Erlbaum, 248 p.
- TALMY, L. 2000. *Toward a cognitive semantics*. Cambridge, MIT Press, vol. 2, 1076 p.
- TRAUGOTT, E.C. 2007. The concepts of constructional mismatch and type-shifting from the perspective of grammaticalization. *Cognitive Linguistics*, 18(4):523-557.

Submissão: 05/08/2009

Aceite: 06/11/2009

**Maria Margarida Martins Salomão**

Universidade Federal de Juiz de Fora

Campus Universitário - Martelos

Rua José Lourenço Kelmer, s/n

36036-900, Juiz de Fora, MG, Brasil