

Um modelo
de restrições
semântico-selecionais
para sistemas
de processamento
automático de
linguagem natural

Luiz Alexandre Mattos do Amaral*

Resumo



Este trabalho apresenta um modelo de restrições semânticas de complementos verbais em português para ser implementado em sistemas de processamento automático de linguagem natural. Esse modelo utiliza informações sobre descrições verbais provenientes de uma gramática de valências para enriquecer as entradas lexicais em HPSG (Head Driven Phrase Structure Grammar).

Palavras-chave: *Linguística Computacional, sintaxe, seleção semântica, HPSG.*

Abstract

This paper presents a semantic selectional restriction model to be implemented in natural language processing systems. This model uses the information about the

* Ohio State University.

descriptions of Portuguese verbs, which were developed based on a model of valence grammar, to enrich verb entries in HPSG (Head Driven Phrase Structure Grammar).

Keywords: *Computational Linguistics, semantic selectional restrictions, HPSG.*

Introdução

A formulação de mecanismos computacionais que lidam com o processamento de linguagem natural apresenta vários níveis de dificuldade relativos aos diferentes aspectos descritivos envolvidos. Descrever uma língua humana a ponto de transformar essa descrição em um algoritmo pode se tornar uma tarefa extremamente complexa devido à grande diversidade de eventos que ocorrem paralelamente quando essa língua está sendo utilizada.

As restrições semânticas impostas por um verbo a seus complementos formam um desses aspectos lingüísticos que apresentam considerável complexidade, quando tentamos descrevê-los de uma forma computacionalmente operável. Porém, devido à sua grande importância em mecanismos automáticos de compreensão de linguagem natural, é impossível esquecê-las quando tentamos solucionar certos problemas típicos de mecanismos computacionais de reconhecimento.

A abordagem aqui proposta lida com essa questão a partir do uso de descrições verbais sugeridas por uma teoria da gramática de valências (GV), apresentada pelo professor Francisco S. Borba (1996), e dos verbetes do Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil (Borba, 1990). Essas descrições verbais são revistas e adaptadas para enriquecer as entradas lexicais de verbos na Head Driven Phrase Structure Grammar (HPSG), de Pollard e Sag (1994).

A HPSG foi escolhida como teoria gramatical de base para nossa descrição já que, além de contar com um desenvolvimento teórico extremamente detalhado, é descritivamente atraente para fins computacionais. Na verdade, sua criação e seu desenvolvimento podem ser atribuídos, em grande parte, a pesquisas na área de Lingüística Computacional.

Desta forma, esse trabalho visa à contribuir para a pesquisa na área de compreensão de linguagem natural através da sugestão de um modelo que aproxima duas teorias distintas, na medida em que utiliza os dados apresentados por uma (GV) para aumentar o poder descritivo da outra (HPSG).

1 - Restrições Semântico-Selecionais

As restrições seccionais referem-se a características semânticas que são impostas aos elementos de uma oração a partir das relações de significado que se constróem no enunciado. Por exemplo, atribuímos às restrições seccionais o fato de (1) ser estranho ao ouvinte e (2) ser perfeitamente aceitável.

(1) ? João bebeu o guarda-chuva.

(2) João bebeu o suco de laranja.

O que está em jogo é o fato de o verbo *beber* normalmente introduzir uma restrição para que seu objeto seja uma entidade do tipo “bebível”. Como

guarda-chuva não é normalmente reconhecido como sendo uma bebida e *suco de laranja* é tido como um líquido ingerível, somente o segundo não causa estranhamento e serve de complemento para o verbo *beber*. Porém, devemos lembrar que as restrições seletivas tentam delimitar aspectos da dependência contextual da interpretação. Assim, não podemos afirmar que exista um erro inerente em (1), uma vez que podemos pensar em contextos onde *guarda-chuva* seja o nome de uma bebida. Androutsopoulos & Dale (1997) nos lembram que “restrições seletivas são tipicamente usadas para capturar fatos sobre o mundo que são geralmente, mas não necessariamente, verdadeiros”. Existem, então, limitações óbvias ao uso de restrições seletivas. Por mais que se descrevam os diferentes sentidos de um mesmo verbo, incluindo usos metafóricos, sempre haverá um novo uso sendo criado ou ainda não descrito. Além disso, temos que lidar com dificuldades apresentadas em construções negativas, como podemos observar no exemplo (3), onde não existe nenhum problema para a interpretação do enunciado.

(3) João não bebeu o guarda-chuva.

Contudo, não podemos evitar de tratar de tais restrições, pois elas são responsáveis em grande parte pelo processo de interpretação de enunciados. Como afirmam Androutsopoulos & Dale (1997),

“Despite their limitations, selectional restrictions have proven very useful in practical applications, and they have been employed in essentially all large-scale natural language understanding systems¹. Apart from blocking pragmatically ill-formed sentences, selectional restrictions are typically used in sense disambiguation, syntactic disambiguation, and anaphora resolution.”

Vejamos em português dois exemplos em que a restrição seletiva tem um papel fundamental na compreensão da sentença. Nos exemplos (4) e (5), vemos que não há ambigüidade quanto ao sentido da palavra *controlador*, uma vez que, em uma leitura comum, o verbo *consertar* pede que seu complemento denote um nome não-animado, fazendo com que o termo *controlador do sistema* seja entendido como uma máquina, enquanto que o verbo *telefonar* seleciona um sujeito mais humano, atribuindo ao *controlador do sistema* o fato de ser uma pessoa.

(4) Fernando consertou o controlador do sistema.

(5) O controlador do sistema telefonou para você ontem.

Já no exemplo (6), a restrição imposta pelo verbo *morar* ao seu sujeito faz com que o pronome *que*, sujeito da oração subordinada, seja compreendido como os imigrantes e não como os países europeus.

¹ Os autores citam em seu artigo dois exemplos de trabalhos que fizeram uso de restrições seletivas em seus desenvolvimentos: (1) Alshawi H, editor. The core Language Engine. MIT. Press, Cambridge, Massachusetts, 1992; (2) Martin P, Appelt D, and Pereira F. Transportability and Generality in a Natural-Language Interface System. In Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, pages 573 - 581. Morgan Kaufmann, 1983.

(6) Os imigrantes dos países europeus que moram há mais de um ano no Brasil podem optar pela cidadania brasileira.

Em nossa abordagem, queremos sugerir um tratamento que lide com as restrições semânticas a partir da descrição verbal feita pela equipe do Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil (Borba 1990), inserindo as informações apresentadas dentro do modelo descritivo da HPSG. O objetivo é propor um uso computacional dos dados extremamente ricos e detalhados que foram compilados pela equipe da UNESP durante a confecção da obra. Esses dados nos permitem montar um modelo que pode ser usado em diferentes sistemas que tenham a necessidade de algum tipo de tratamento das informações semânticas dos complementos verbais. Como veremos a seguir, a informação disponibilizada no Dicionário nos permite diferenciar os principais usos dos verbos no português do Brasil, além de trazer dados sobre as valências quantitativa, qualitativa e semântica e as propriedades sintático-semânticas dos complementos verbais. Todos esses dados são apresentados a partir de uma classificação verbal e diretamente relacionados com os diversos significados de cada verbo.

2 - Gramática de Valências

O conceito de valência tem sido usado por diversos estudiosos na tentativa de criar mecanismos descritivos capazes de capturar certos fenômenos recorrentes na formulação de enunciados, cujas explicações não parecem ser possíveis apenas por critérios sintáticos ou semânticos analisados de forma separada. Desde a primeira proposta de Tesnière (1965) até os estudos de Busse & Vilela (1986) para o português de Portugal e Borba (1990; 1996) para o português do Brasil, as diferentes definições do conceito de valência giram em torno da capacidade que os itens lexicais possuem de estabelecer determinadas relações entre si a partir de seu significado. A abrangência das relações estabelecidas e os critérios utilizados para determiná-las vão variar de proposta para proposta, porém a rede de relações entre as palavras de um enunciado vai ser o alvo principal das descrições valenciais.

A proposta de Borba (op. cit.) parte do mesmo princípio apresentado por Tesnière (1965), no qual o núcleo oracional é composto pelo verbo que estabelece o número de actantes por ele regidos, determinando, assim, sua valência. Além disso, por ser o núcleo oracional e manter relações de dependência com seus actantes, o verbo é o elemento principal na formação de um “esquema abstrato” de modelos oracionais da língua.

Borba amplia a noção inicial de valência ao analisar as relações de dependência entre outras classes de elementos que não estejam diretamente ligadas ao verbo. Para ele, o estudo do regime dos nomes, adjetivos e alguns advérbios se baseia na capacidade desses elementos de se ligarem a outros elementos oracionais, e possibilita a delimitação de suas respectivas classes. Ele propõe, então, o uso do termo valência em três níveis: (1) a valência quantitativa está relacionada com o número de argumentos que um predicado pode ter. Assim, um item lexical pode ser avalente, monovalente (P(A)), divalente (P(A1, A2)), trivalente (P(A1,A2,A3)) ou

tetravalente (P(A1,A2,A3,A4)); (2) a valência qualitativa ou morfossintática descreve as características das classes dos itens que podem preencher um determinado argumento, assim como as propriedades morfológicas desses itens; (3) a valência semântica, por sua vez, determina as funções (papéis) temáticas dos argumentos, as restrições seletivas dos itens que compõem os argumentos e os traços que definem as características de cada categoria (+- animado, +- humano etc)².

A análise dos elementos estruturais, a partir da teoria aqui exposta, conta ainda com observações a respeito das relações sintático-semânticas da estrutura conceitual do léxico. Essa análise enfoca um aspecto do significado dos enunciados que se encontra atrelado às estruturas conceituais projetadas a partir de regras de inferência e de pragmática. O relacionamento de tais estruturas envolve constituintes conceituais e constituintes sintáticos dentro de uma abordagem onde certas categorias sintáticas podem expressar determinadas categorias conceituais. Assim, um sintagma nominal pode expressar categorias conceituais como objeto, evento, qualidade etc. Observando a categoria conceitual expressa por um sintagma nominal (adjetival ou preposicional) e a sua relação com outras estruturas conceituais dentro do enunciado, podemos determinar os papéis temáticos do sintagma em questão. Os papéis temáticos, por serem noções relacionais, só podem ser inferidos se analisarmos as configurações estruturais envolvidas no enunciado.

Uma maneira de representarmos os papéis temáticos é fazendo uso de uma gramática de casos, em que um caso apontará a atuação dos argumentos na predicação. O valor semântico de cada argumento vai determinar a nomenclatura utilizada na classificação dos casos. Esse valor semântico não está ligado a nenhum item lexical específico, só podendo ser determinado a partir da relação entre valores sintático-semânticos.

Para escolher os casos que serão utilizados em sua teoria, Borba (1996) segue Fillmore e trabalha com os seguintes casos: agente (Ag), beneficiário (B), causativo (C), comitativo (Co), experimentador (Ex), instrumental (I), locativo (L), meta (M), objetivo (Obj), origem (Or), resultativo (R) e temporal (Tp).

A maneira como os papéis temáticos são estabelecidos dentro da teoria apresenta logo de início algumas vantagens para o uso computacional. A primeira delas é a de que eles nunca são atribuídos negativamente, ou seja, é sempre necessário que o sintagma corresponda a uma posição argumental específica dentro da estrutura conceitual, o que faz com que possamos lhe atribuir papéis semânticos capazes de entrar em um esquema de hierarquia semântica. A segunda vantagem é que cada predicado possui apenas um tipo de relação casual, facilitando a diferenciação entre eles.

3 - Matriz Valencial

Durante a década de 80, um grupo de pesquisadores da UNESP desenvolveu uma exaustiva pesquisa sobre os verbos da língua portuguesa e

²

A valência semântica tem um papel de destaque no presente trabalho, cujo principal objetivo pode ser resumido como sendo o de avaliar as informações apresentadas por Borba sobre a valência semântica dos verbos em português, a partir do seu emprego em um mecanismo computacional de restrições seletivas dos complementos verbais.

como fruto desse trabalho publicaram uma belíssima obra intitulada Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil (Borba, 1990). Nela, os verbos são apresentados a partir de diversos critérios descritivos que nos permitem observar, entre outras coisas, os três níveis valenciais descritos anteriormente. Assim, é possível selecionar informações nos verbetes para compor uma matriz das características verbais relevantes para a descrição almejada.

Em nosso trabalho, utilizamos algumas das informações presentes para criar uma tabela que é usada como base do modelo descritivo dos diagramas AVM (attribute-value matrix), como veremos mais à frente. O exemplo apresentado na Figura 1 é o do verbo *comer*.

Classe	Valor (Definição)	SUJEITO		COMPLEMENTO		
		Caso	Descrição Semântica	Caso	Descrição Semântica	Desc. Sint.
Ação	Engolir para se alimentar; ingerir.	Ag	+animado	Obj	Alimento sólido	SN
Ação-processo	Corroer; destruir	C	-animado elemento corrosivo; doença	Obj	-animado	SN
Ação-processo	Gastar; dilapidar	Ag	nome	Obj	Tempo; bens; acontecimento	SN
Ação-processo	Suprimir; omitir	Ag	+animado	Obj	Elementos lingüísticos	SN
Ação-processo	Eliminar	Ag	+animado	Obj	Peças de jogo de tabuleiro	SN
Ação	Provar; experimentar	Ag	+animado	Obj	Alimento sólido	SP

Figura 1

O quadro segue exatamente as premissas das descrições apresentadas na teoria de valências. Assim, a primeira coluna apresenta a classe dos verbos de acordo com cada uso. Vale lembrar que as classes só podem ser atribuídas dentro de um contexto, pois um mesmo verbo pode ter mais de uma classe, como é o caso do verbo *comer*. A segunda coluna traz o que denominamos neste trabalho o valor do verbo, ou seja, em uma situação em que seus complementos tenham determinadas características, o verbo terá um determinado sentido (definição; valor). Normalmente, esse valor pode ser atribuído a partir de sinônimos. A terceira coluna descreve as informações sobre o sujeito e se divide em duas outras colunas, uma com o caso e outra com as características semânticas que o sujeito deve apresentar. O mesmo ocorre com a coluna seguinte, que é dividida em três, trazendo não só as duas informações já citadas, mas também as características sintáticas do

complemento. Essa última informação não é necessária para o sujeito, uma vez que não lidaremos com outro tipo de sujeito que não seja um SN.

O interessante a respeito de uma descrição dessa natureza é o fato de os valores estarem todos atrelados uns aos outros, criando uma rede de dependência descritiva que só se estabelece quando os valores necessários são determinados e confrontados. Por exemplo, *comer* só será um verbo que indique ação-processo e sinônimo de *corroer* se o seu sujeito for -animado e representar um elemento corrosivo ou doença, e seu complemento for -animado, fazendo com que os casos atribuídos sejam, respectivamente, causativo e objetivo.

Existem, porém, ocasiões em que um dos valores é perfeitamente dispensável, pois toda a dependência descritiva encontra-se em um único elemento, como na terceira definição da tabela, em que o complemento vai determinar sozinho o valor do verbo. Esse fator não vai dificultar em nada a transcrição que faremos a seguir para a HPSG.

4 - A Head-Driven Phrase Structure Grammar

A HPSG (Head-Driven Phrase Structure Grammar) é uma teoria gramatical que foi desenvolvida nos anos 80 por dois pesquisadores norte-americanos e que ganhou enorme notoriedade em diversos departamentos de lingüística devido às suas características descritivas e à sua abrangência teórica. Desde suas primeiras formulações, a HPSG tem sido utilizada por pesquisadores no desenvolvimento de implementações computacionais³. Atualmente, a HPSG é um dos formalismos gramaticais mais utilizados nos Estados Unidos, Canadá, Europa e Austrália, além de contar com diversos estudos desenvolvidos por centros de pesquisa asiáticos no Japão, na Coreia e em Taiwan.

Algumas noções básicas da HPSG podem ser resumidas da seguinte maneira:

- 1 - A HPSG trabalha com a noção de lexicalismo estrito. Os princípios que regem a formulação dos itens lexicais não são os mesmos que licenciam as construções frasais.
- 2 - As estruturas são analisadas na sua forma superficial. Evita-se a criação de estruturas abstratas para a explicação de fenômenos sintáticos ou semânticos.
- 3 - A organização lingüística é feita de forma hierárquica.
- 4 - Dentro da hierarquia da HPSG, os núcleos lexicais são responsáveis pela seleção de seus complementos, sujeitos e especificadores. A atribuição de papéis semânticos, casos e concordância em gênero e número é feita segundo a arquitetura da estrutura de atributos que reflete a valência do item lexical.
- 5 - O léxico é responsável por um fenômeno chamado de "classificação múltipla", ou seja, a informação lexical não é simplesmente listada; ela é organizada em termos de heranças múltiplas e regras léxicas

³ Um grande projeto patrocinado pelos laboratórios da Hewlett-Packard, entre 1980 e 1991, impulsionou os estudos da HPSG. Nele trabalharam, além dos autores da teoria, uma série de alunos e pesquisadores de instituições americanas e européias.

que permitem que as propriedades morfosintáticas das palavras sejam derivadas a partir da lógica presente na descrição lexical.

- 6 - Tipos frasais também são tratados a partir de mecanismos de heranças múltiplas, o que permite a formação de generalizações sobre diversos tipos de construções.

O principal mecanismo descritivo utilizado pela HPSG chama-se Estrutura de Atributos⁴. Uma Estrutura de Atributos (EA) é uma maneira de representar informações gramaticais que consiste em especificar uma lista de características estruturais, cada qual tendo um valor determinado⁵. Quando organizamos uma classificação lexical, normalmente observamos a parte do discurso em que o item escolhido se insere e então lhe atribuímos uma classe. Ao usarmos uma EA, ao invés de associarmos um item lexical a uma categoria, descrevemos um item, uma categoria ou até mesmo uma estrutura como um grupo complexo de propriedades gramaticais. Essa descrição é composta de duas partes: o atributo a ser descrito (que aparece em letra maiúscula) e o valor referente a ele (em letra minúscula). Além disso, é importante notarmos que o valor conferido a um atributo pode ser um item ou uma outra EA. Para que a descrição de uma EA fique mais clara, usaremos a notação de chaves⁶, como na Figura 2:



Figura 2

A elaboração de uma estrutura de atributos é feita a partir da composição de uma teoria que preveja os tipos de atributos necessários e sua articulação com níveis de predicação. O julgamento das orações bem formadas de uma determinada língua é feito em comparação com as estruturas de predicação previstas, e o que se espera dessas estruturas é que elas tenham um nível de adequabilidade descritiva satisfatório. Assim, podemos formular uma teoria lingüística em termos de restrições que regem as categorias de uma determinada estrutura de atributos.

Neste trabalho, vamos apresentar as descrições verbais apenas com os atributos essenciais para a compreensão de nossa proposta⁷. Assim, uma descrição verbal básica contará com os atributos CATEGORY (CAT), CONTENT (CONT) e CONTEXT/BACKGROUND (CX/BG).

⁴ Optamos por essa tradução do original FEATURE STRUCTURES por acharmos que ela expressa melhor o conceito em questão.

⁵ O uso de EAs em HPSG demonstra uma opção por uma arquitetura que valoriza a riqueza de informações em um mesmo item. Isto é, ao contrário de outros formalismos como a LFG, que contém diferentes estruturas para diferentes características, a HPSG disponibiliza informações fonéticas, sintáticas, semânticas e pragmáticas de um determinado item na mesma EA.

⁶ Este tipo de notação é a base dos diagramas AVM (attribute-value matrix) usados nas descrições em HPSG.

⁷ O leitor interessado em conhecer todos os atributos e valores previstos para a categoria verbo na HPSG poderá encontrá-los em Pollard & Sag (1994).

5 - Descrição Verbal

Utilizando o exemplo da Figura 1, podemos demonstrar como as informações previstas pela descrição valencial podem ser incorporadas à descrição verbal em HPSG.

Vamos tomar como definição de base o primeiro valor do verbo *comer* apresentado na tabela. Esse valor tem a seguinte configuração: o verbo *comer* indica ação, possui sujeito agente e +animado, complemento objetivo designativo de alimento sólido, e seu significado é “engolir para se alimentar, depois de ter mastigado, ingerir”, como no exemplo: *Paula comeu o bolo*. Assim, a descrição do verbo *comer* com esse valor fica como na Figura 3.

Os dois primeiros níveis valenciais (o da valência quantitativa e o da valência morfossintática) não apresentam nenhuma dificuldade na aproximação entre as duas teorias, uma vez que o seu posicionamento descritivo⁸ é facilmente observável na HPSG. Os dados referentes ao número de complementos de um verbo e as características morfossintáticas desses complementos serão apresentados na EA CAT.

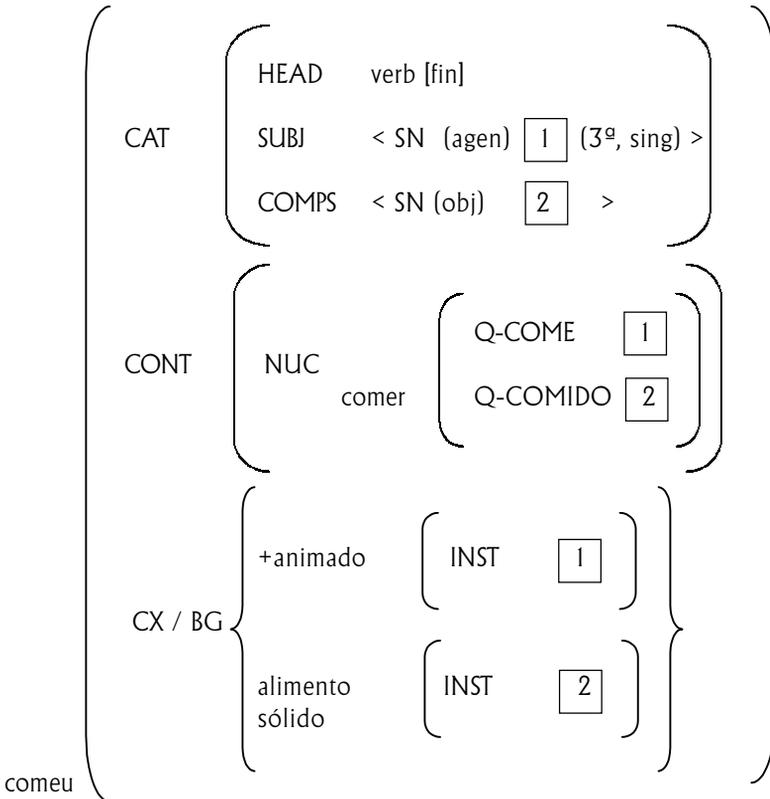


Figura 3 - Engolir para se alimentar; ingerir.

⁸ Chamamos de posicionamento descritivo o local dentro do diagrama AVM onde se pode encaixar uma determinada informação.

A descrição da valência semântica será feita em duas EAs distintas, cada qual tendo uma função específica dentro da arquitetura da HPSG. A primeira (CONT) especifica o valor semântico das estruturas sintáticas descritas em CAT. Assim, o SN sujeito receberá a classificação Q-COME (quem come, agente da ação) e o SN complemento será Q-COMIDO (que é comido, objeto da ação). Esses atributos são marcados por um índice referencial que faz com que suas características sejam estruturalmente compartilhadas com todos os atributos e/ou valores que tenham o mesmo índice referencial. Já a segunda EA (CX/BG) é responsável pela restrição semântica dos actantes do verbo. Isso é possível graças ao papel atribuído a essa EA, que é descrita em Pollard & Sag (1994) como responsável pela escolha das condições contextuais essenciais para a boa formação do enunciado. Essa característica descritiva se baseia em um princípio da HPSG chamado Princípio de Consistência Contextual, que especifica que o valor de BG de cada enunciado é determinado pela união dos valores de BG do núcleo de DTRS (atributos filhas). Desta forma, o valor do verbo *comer* escolhido só será atribuído ao verbo em questão se o nome que forma o núcleo do SN actante tiver o mesmo traço semântico.

Uma vez definido o valor padrão para o verbo, podemos construir todas as outras definições, ressaltando as diferenças que serão responsáveis pela atribuição de novos significados.

A segunda definição vai apontar para o verbo *comer* indicando ação-processo, com sujeito causativo expresso por nome não-animado designativo de elemento corrosivo e doença, e com complemento objetivo também expresso por nome não-animado. O verbo *comer* com tal valência vai ser sinônimo de *corroer*, *destruir*, como no exemplo: *a ferrugem comeu as peças de ferro*. Para a segunda definição, existem modificações a serem feitas em CAT, em que o sujeito ganha o caso causativo, e em CX/BG, em que os valores contextuais serão diferentes. A Figura 4 ilustra essas modificações.

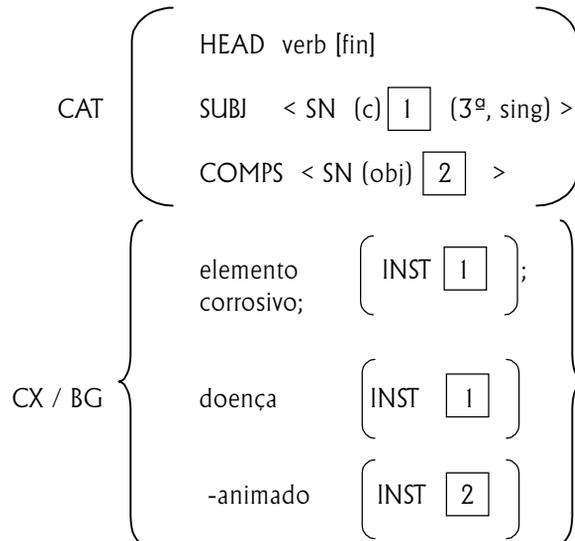


Figura 4 - Corroer; destruir.

As definições 3, 4 e 5 vão se basear no verbo *comer* indicando ação-processo com sujeito agente e com todos os complementos expressando o caso objetivo. Dessa forma, o que na verdade vai diferenciar cada significado do verbo será o valor semântico dos complementos. Na terceira definição, o complemento do verbo expressa nomes designativos de bens, de tempo ou de acontecimento. Seu significado é *gastar*, *dilapidar*, como em: *as minhas dívidas comeram toda a fortuna*. O valor de CX/BG ficará como na Figura 5. Todos os valores têm o mesmo índice porque o sujeito não conta com nenhuma restrição semântica especial.

O ponto e vírgula vai marcar as diferentes possibilidades de valor que o atributo pode ter, ou seja, esse significado pode ser atribuído ao verbo *comer* se o seu complemento tiver qualquer um dos três traços utilizados para descrever os tipos de EA que servem de valor para CX/BG na Figura 5.

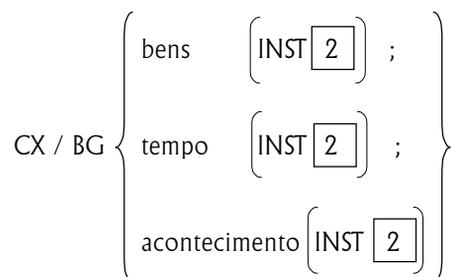


Figura 5 - Gastar; dilapidar.

A quarta definição descreve *comer* como um sinônimo de *suprimir*, *omitir*, e seu complemento designa elementos lingüísticos, como no exemplo: *você comeu algumas palavras no texto*. As alterações se apresentam na Figura 6.

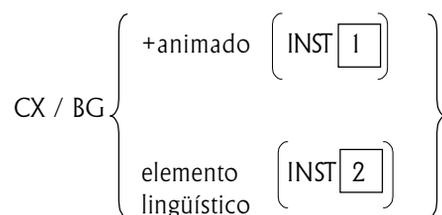


Figura 6 - Suprimir; omitir.

A quinta e última definição deste grupo apresenta o complemento do verbo *comer* como sendo designativo de peças de jogo de tabuleiro, como damas e xadrez. Seu significado é sinônimo de *eliminar*, como podemos ver nos exemplos: *se eu comer o seu rei acaba o jogo*, *ele comeu três peças com a dama*. Sua descrição está apresentada na Figura 7.



Figura 7 - Eliminar

O sexto, e último, uso foi escolhido porque é o único no qual o valor de *comer* é marcado por um traço na valência sintática. O verbo *comer* terá o sentido de *provar* quando tiver as mesmas características da primeira definição apresentada, porém com um SP como complemento no lugar de um SN, como nos enunciados: *a Clara já comeu desse doce?*, *Eu nunca comi da sua mousse*. O fator que marca a sua distinção da descrição padrão não se encontra em BG, uma vez que as características semânticas do sujeito e do complemento são as mesmas. Devemos, então, reformular o valor de CAT para que, ao invés de um SN, seja encontrado um SP como complemento do verbo, como mostra a Figura 8.

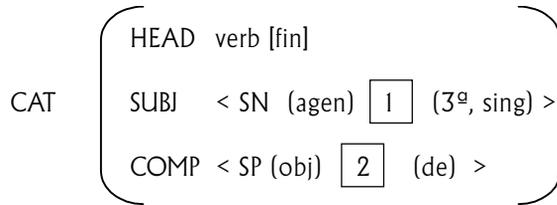


Figura 8 - Provar; experimentar.

6 - Problemas de Processamento e Contra-Exemplos

Como explicitado anteriormente, existem problemas impostos ao processamento de certos enunciados que fogem à abordagem aqui apresentada ou que demonstram alguma dificuldade específica de tratamento a partir da matriz descrita, gerando contra-exemplos que só poderão ser tratados a partir de um refinamento nos mecanismos descritivos.

Várias são as razões para a existência desses contra-exemplos. A primeira delas é a imprecisão da escolha de critérios para a criação das subcategorias semânticas, que pode possibilitar diversas falhas na pré-seleção de itens para o preenchimento dos espaços dos actantes. Isso ocorre pois, além dos traços semânticos necessários para que um elemento funcione como sujeito ou complemento de um verbo, existem questões de uso específico que vetam certas construções. É o caso dos exemplos em (7), em que usamos a terceira definição do verbo *comer* (*gastar, dilapidar*) e os dois complementos se referem a bens.

- (7) a) Meus filhos comeram todo o meu dinheiro esse mês.
b) ? Maria comeu minha fazenda esse mês.

Fora de um contexto, a segunda frase causa estranheza, porque seu uso não é corriqueiro e o falante pode não atribuir o valor correto ao verbo *comer*. Porém, nos parece que mesmo usos menos freqüentes de verbos, quando contextualizados, não causam tantos problemas de compreensão.

A segunda razão decorre de um fenômeno de relação entre os actantes do verbo que foge ao mero estabelecimento de regras semânticas de predicação, como a existência de uma relação de significados entre o sujeito e o complemento que impede a atribuição de determinado valor ao verbo mesmo que as características semânticas estejam presentes, causando estranhamento ao ouvinte. Vamos utilizar como exemplo a segunda definição do verbo *comer* (*corroer; destruir*). Segundo a descrição, se o sujeito for um sub-item de doença e o complemento algo inanimado, a frase será aceita. Com certeza, o leitor pode pensar em diversos exemplos em que uma doença não vai corroer ou destruir algo inanimado como uma mesa. Porém, podemos ir além, descrevendo o complemento do verbo como órgãos ou partes do corpo para a definição em que o sujeito seja uma doença, e mesmo assim teríamos problemas, pois características semânticas presentes nos itens de doença criariam restrições quanto aos complementos aceitos. Vejamos os exemplos em (8).

- (8) a) A hepatite comeu seu fígado.
b) ? A hepatite comeu seus pulmões.
c) A tuberculose comeu seus pulmões.
d) ? A tuberculose comeu seu fígado.

Os exemplos (b) e (d) causam estranheza ao interlocutor devido à relação de dependência semântica que existe entre o sujeito e o complemento. Como a hepatite é uma doença do fígado e a tuberculose uma doença dos pulmões, qualquer outro órgão que apareça na posição de complemento do verbo vai causar problemas de compreensão ao interlocutor. Como esse, existem outros casos em que a classificação genérica de um subgrupo não vai prevenir a aparição de contra-exemplos.

Uma terceira razão, talvez a mais importante de todas, para a aparição de contra-exemplos é o fato de a língua estar sempre em transformação, o que faz com que usos que não eram aceitos anteriormente pelos falantes passem a ser aceitos, como no caso da aparição de novas metáforas. Assim, torna-se inviável uma descrição do papel semântico na predicação que dê conta de todos os exemplos de uma língua.

Conclusão

Este trabalho apresentou uma maneira de implementar restrições semânticas para complementos verbais em um sistema computacional, tendo como base os dados disponíveis no Dicionário Gramatical de Verbos do

Português Contemporâneo do Brasil (Borba, 1990). Além da necessidade de aprimoramento de mecanismos computacionais que sejam capazes de lidar com o problema das restrições semânticas, existem outros aspectos que destacam a relevância da presente abordagem.

Primeiro, poucos trabalhos que lidam com HPSG têm sido desenvolvidos visando ao português, enquanto que para outras línguas (não necessariamente o inglês), as pesquisas já se encontram bem mais avançadas. O reflexo desse fenômeno pode ser observado nas últimas conferências internacionais da HPSG, onde pouquíssimos trabalhos que tratam de nossa língua foram apresentados e os que estiveram presentes eram em sua totalidade representantes do português europeu.

Segundo, uma grande virtude da abordagem desenvolvida é o fato de ela lidar com uma fonte (o Dicionário) onde os dados foram cuidadosamente levantados e apresentados após um longo trabalho de pesquisa de campo, no qual foram recolhidas amostras lingüísticas do português do Brasil e todas as informações foram dispostas com base em uma teoria de gramática já desenvolvida. Assim, quando usamos esses dados para compor uma descrição computacional, estamos descrevendo o uso verbal feito por falantes contemporâneos da língua portuguesa de nosso país, o que enriquece o resultado obtido.

Terceiro, a abordagem aqui apresentada propõe a junção de dados referentes a duas teorias para a resolução de um difícil problema de processamento. Com isso, ela reforça a opinião daqueles que acreditam ser possível a investigação de métodos que aproveitem diversas abordagens para a solução de problemas. De uma forma mais ampla, essa talvez seja uma das contribuições mais importantes que a área computacional tem dado aos estudos lingüísticos, já que, nos estudos em lingüística computacional, existe uma grande motivação para propor soluções implementáveis, não importando a origem teórica do tratamento.

Bibliografia

ANDROUTSOPOULUS, I. & DALE, R. *Semantic Selectional Restrictions in HPSG*. Technical report C/TR, Sydney: Macquarie University, 1997.

BORBA, F. S. *Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil*. São Paulo: Edunesp, 1990.

_____. *Uma Gramática de Valências para o Português*. São Paulo: Ed. Ática, 1996.

BUSSE, W. & VILELA, M. *Gramática de Valências: Esboço de Apresentação e Aplicação ao Português*. Coimbra: Ed. Almedina, 1986.

COPESTAKE, A. *The Representation of Lexical Semantic Information*. PhD thesis, University of Sussex, 1992.

DAVIS, T. *Lexical Semantics and Linking in the Hierarchical Lexicon*. PhD thesis, Stanford University, 1996.

KILGARRIFF, A. *Inheriting Verb Alternations*. Proc 6th European ACL pp 213-221. Utrecht, Holland. 1993.

- _____. *Inheriting Polysemy. Computational Lexical Semantics*. Patrick St. Dizier & Evelyne Viegas (Eds.) p. 319-335 CUP, 1995.
- POLLARD, C. & SAG, I. A. *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press, 1994.
- RADFORD, A. *Syntactic Theory and the Structure of English*. Cambridge: Cambridge University Press, 1997.
- SAG, I. & WASON, T. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, 1999.
- TESNIÈRE, L. *Éléments de Syntaxe structurale*. Paris: Klincksieck, 1965.
- VILELA, M. *Gramática de Valências: Teoria e Aplicação*. Coimbra: Livraria Almedina, 1992.
- _____. *Gramática da Língua Portuguesa*. Coimbra: Livraria Almedina, 1999.

Um modelo
de restrições
semântico-
selecionais
para sistemas
de processamento
automático de
linguagem natural

Luiz Alexandre
Mattos do Amaral

