

## 4 Estudo de caso

Esse capítulo demonstrará o levantamento e a atribuição de conceitos e estruturas de conceitos que representarão os significados de entradas de um dicionário eletrônico. Conjuntamente à estrutura de conceitos dessas entradas também serão demonstradas algumas estruturas de argumentos. Será simulado aqui o trabalho manual de executar essa tarefa, para o leitor se familiarizar com o modelo. No entanto, por acreditar que o levantamento das estruturas de conceitos é algo mecânico, apenas dependendo de um cotejo entre as paráfrases, para usos reais esse trabalho deverá ser automatizado.

Com a automatização do processo de atribuir estruturas de conceitos às entradas, o trabalho do lingüista passaria a se limitar ao reconhecimento das paráfrases existentes no uso analisado para cada entrada, à atribuição das classes de palavras de cada entrada e ao levantamento dos argumentos de cada entrada. O desenho final da estrutura de conceitos, a distribuição dos conceitos entre as paráfrases, o relacionamento entre conceitos e a própria criação de conceitos são tarefas possíveis e desejáveis de serem automatizadas. Criar conceitos e estruturas de conceitos não é uma tarefa difícil, apenas é trabalhosa porque envolve muitos dados – como veremos –, no entanto, é tarefa possível para a computação.

Para que fosse simulada a atribuição de estrutura de conceitos, primeiramente foi levantado um corpus. O corpus é necessário por várias razões. A primeira razão é definir sobre qual realidade lingüística particular o dicionário recai. Isso é muito importante, uma vez que o presente modelo deverá ter maior eficácia em usos restritos. Quanto mais restrita for a realidade lingüística maior será o número de paráfrases encontradas.

Isso pode ser demonstrado num exemplo simples dentro do uso escolhido para esse corpus, o jornalismo esportivo. Provavelmente, em contexto lato, não seria concebível que “contratar alguém para trabalhar num lugar” pudesse ser parafraseado por “alguém ir para um lugar”. Mas, no contexto específico tratado, isso é plenamente

concebível. Vejamos exemplos retirados de matérias de jornal que divulgam a mesma notícia:

“Falcão assina contrato com o São Paulo”.

“Tricolor contrata Falcão”.

“Falcão vai jogar no São Paulo”.

“O ala Falcão acertou sua ida para o clube do Morumbi”.

“Falcão vai para o São Paulo”.

A notícia é a mesma, a informação também é a mesma, mas são usados verbos absolutamente improváveis como paráfrases. Nesse contexto específico, a informação de um clube contratar um novo jogador pode ser expressa tanto pela construção com o verbo “contratar” quanto com o verbo “ir”. Isso demonstra o quanto as paráfrases são dependentes da especificidade do uso que se faz dos textos.

Essa é a primeira razão para levantar um corpus. A segunda razão é que o processo de atribuição de uma estrutura de conceitos para uma palavra é dependente de todas as paráfrases que essa palavra pode fazer no uso lingüístico trabalhado. Precisa-se do corpus para limitar quais as paráfrases reais são efetivamente realizadas nesse uso.

O corpus usado é bem reduzido para que seja possível mostrar passo a passo, no espaço de um capítulo, como executar a representação da semântica e da gramática neste modelo. O corpus é composto de oito pequenos textos retirados de matérias reais de jornais diferentes que relatam a mesma notícia: um Fla-Flu que terminou em empate. Dessas matérias foram removidas para compor o corpus apenas as frases que continham pelo menos uma destas cinco idéias: jogador, time, jogar, empatar e fazer gol<sup>1</sup>.

Em seguida serão apresentados os processos de levantar e atribuir os conceitos das palavras de um corpus.

---

<sup>1</sup> O corpus completo está no anexo no final dessa dissertação.

#### 4.1 O dicionário eletrônico

O processo básico de levantar as estruturas de conceitos deve levar em conta somente as paráfrases apresentadas no corpus, como se existissem apenas as formas ali apresentadas para expressar as informações<sup>2</sup>. Mas se tentou, na medida do possível, ao levantar as estruturas de conceitos, prever o máximo de construções semelhantes que poderiam aparecer no contexto do jornalismo esportivo. Isso fez com que, apesar de pequenos, os exemplos de atribuição de estruturas de conceitos demonstrassem uma complexidade lingüística (e de certo modo pragmática) próxima à real.

O primeiro passo é definir, dentre as cinco idéias expressas, qual é a mais simples, isso é, qual das idéias não é composta por nenhuma outra idéia expressa.

A informação contida na expressão “fazer gol”<sup>3</sup> pressupõe um jogador ou um time que a realize. O mesmo se pode dizer de “jogar” e “empatar”. É difícil definir entre “jogador” e “time” qual é a idéia mais simples, pois parecem se relacionar uma com a outra. Portanto, vamos trabalhar com as duas.

Sabemos pragmaticamente que um time é um conjunto de jogadores, no entanto o processo de atribuição de uma estrutura de conceitos é algo mecânico, e não exatamente pragmático. Para saber se entre as palavras ou expressões que contenham as idéias de “time” e de “jogador” há compartilhamento de conceitos, precisamos saber como essas palavras ou expressões se comportam como paráfrases.

No corpus, temos as seguintes palavras e expressões que expressam a idéia de “time”, ou em que, no que expressam, está contida a idéia de time: “equipe da Gávea”, “equipe flamenguista”, “ambos os lados”, “as duas equipes”, “derrota rubro-negra”, “tricolor”, “equipe”, “equipe tricolor”, “Fla”, “Fla-Flu”, “Flamengo”, “Flu”,

---

<sup>2</sup> Por ser um trabalho inadequado às capacidades humanas, é claro que, num uso lingüístico real, as paráfrases não devem ser retiradas de um corpus uma a uma. O julgamento da presença ou ausência de uma paráfrase (bem como de um significado de uma palavra) numa realidade de uso da língua deve ser feito pelo conhecimento lingüístico de quem se propuser a representar as paráfrases. O grau de acerto do buscador orientado a idéia, portanto, será determinado pela falibilidade humana, é o preço a pagar.

<sup>3</sup> Como um significado se define a partir das paráfrases que podem expressá-lo, sempre que nos referirmos a uma idéia, informação etc. o faremos através de uma palavra ou de uma expressão usadas para a referirem. Isso quer dizer que quando for dito algo como “a idéia de ‘time’”, leia-se “a idéia da palavra ‘time’”. Dessa maneira, o significado também virá escrito entre aspas.

“Fluminense”, “gol tricolor”, “clube da Gávea”, “rubro-negro”, “time da Gávea”, “jogadores do Flamengo”, “tricolores”, “time das Laranjeiras”, “time rubro-negro”, “time tricolor”, “time”.

A idéia de “jogador” está contida em: “defesa rubro-negra”, “zaga rubro-negra”, “André”, “atacante”, “atacante rubro-negro”, “atleta”, “centroavante”, “Diego”, “dois jogadores rubros-negros”, “Fellype Gabriel”, “goleiro Kleber”, “Jônatas”, “Leandro”, “Lino”, “goleiro tricolor”, “gringo”, “jogador”, “meia”, “Obina”, “Pet”, “Petkovic”, “Preto”, “Preto Casagrande”, “Renato”, “Rodrigo Tiui”, “Romeu”, “sérvio”, “Souza”, “Tuta”.

Prosseguindo, devemos emparelhar, para cada idéia, as palavras que estabeleçam algum tipo de relacionamento de paráfrases. São esses os tipos de relacionamentos: palavras ou expressões que se substituam, palavras ou expressões que compartilhem parte do significado e palavras ou expressões que contenham todo o significado de outras palavras ou expressões. Dessa maneira podemos listar:

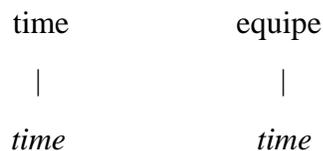
– Palavras ou expressões que se substituam: “equipe da Gávea”, “equipe flamenguista”, “Fla”, “Flamengo”, “clube da Gávea”, “rubro-negro”, “time da Gávea”, “jogadores do Flamengo”, “time rubro-negro”; “tricolor”, “equipe tricolor”, “Flu”, “Fluminense”, “tricolores”, “time das Laranjeiras”, “time tricolor”; “ambos os lados”, “as duas equipes”; “equipe”, “time”.

– Palavras ou expressões que contenham todo o significado de outras palavras ou expressões: a idéia expressa em “equipe” e “time” está contida em “equipe da Gávea”, “equipe flamenguista”, “Fla”, “Flamengo”, “clube da Gávea”, “rubro-negro”, “time da Gávea”, “jogadores do Flamengo”, “time rubro-negro”, “tricolor”, “equipe tricolor”, “Flu”, “Fluminense”, “tricolores”, “time das Laranjeiras”, “time tricolor”.

– Todas as palavras e expressões compartilham parte do significado, o significado que é expresso por “time”.

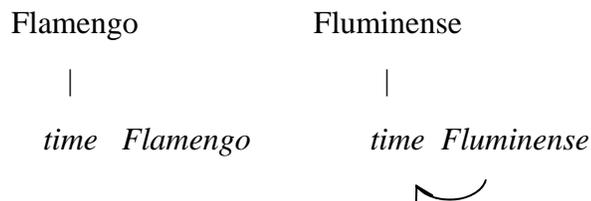
A partir dessas divisões vamos definir as estruturas de conceitos. Se “equipe” e “time” são menos complexos que “Flamengo”, “Fluminense” etc., então devemos

começar por essas palavras. Dessa maneira, atribuiremos a elas o mesmo conceito, representando o significado de ambas as palavras<sup>4</sup>:



**Figura 1**

A palavra “Flamengo” expressa a mesma idéia de “time”, no entanto, essa idéia é especificada. Isso quer dizer que o conceito que representa o significado de “time” deve ser especificado por outro conceito ligado a ele. O mesmo ocorre com “Fluminense”, mas “Flamengo” e “Fluminense” não são especificados da mesma maneira, pois não são paráfrases, portanto os conceitos que especificarão o conceito *time* deverão ser distintos.



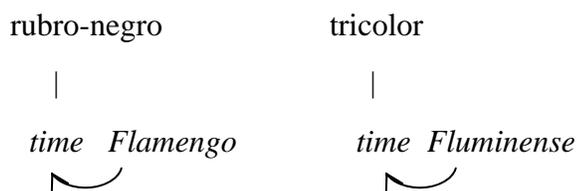
**Figura 2**

Nosso processo sempre começa com as palavras que representam idéias mais simples, isso é, que contenham menos idéias que as outras palavras analisadas, e sempre escolhemos começar pelas palavras simples e não pelas expressões

---

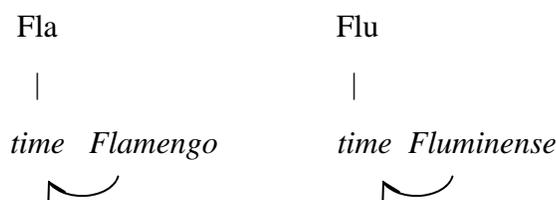
<sup>4</sup> É muito importante notar que a escolha das palavras usadas para nomear um conceito é arbitrária, poderia ser uma letra, como foi feito anteriormente, poderia ser uma seqüência aleatória de caracteres ou números, poderia ser um símbolo, um desenho ou uma cor, poderia ser ainda palavras sem nenhuma ligação com os significados (como o conceito *chuchu*, representando a idéia de “time”). O que importa é que sempre que se esteja referindo a uma determinada idéia, esta seja representada pelo mesmo conceito ou estrutura de conceitos. Escolheu-se para nomear os conceitos palavras que tivessem alguma ligação com os significados por uma necessidade mnemônica. Para diferenciar o conceito *time* da palavra “time”, o conceito vem escrito em itálico e a palavra entre aspas. Escolhemos arbitrariamente o conceito *time* para representar a idéia expressa pela palavra “time”, segundo nossos critérios, poderia ser *equipe*, por exemplo. Mas, após feita a escolha por *time*, todas expressões que contenham essa idéia devem possuir em suas estruturas de conceitos o conceito *time* e não outro.

compostas. Assim sendo, as próximas palavras a receber estruturas de conceitos são os epítetos “tricolor” e “rubro-negro”. Essas palavras agem como sinônimos respectivamente de “Fluminense” e “Flamengo”. Então receberão a mesma estrutura de conceitos que essas palavras.



**Figura 3**

O mesmo acontece com “Fla” e “Flu”:



**Figura 4**

É muito interessante notar as peculiaridades geradas pelas premissas desse modelo. Como um conceito se relaciona a outro pela especificação, acabamos não tendo uma estrutura hierarquizada entre os conceitos *time* e *Fluminense*, como uma interpretação pragmática calmamente estratificaria. Neste modelo, nunca poderá existir hierarquia entre os conceitos por dois motivos. O primeiro é que se quer aproveitar as habilidades do grafo<sup>5</sup> (como a busca de subgrafos em grafos) no nosso modelo, e num grafo não há hierarquia entre seus elementos. A árvore, uma técnica

<sup>5</sup> Lembre-se que numa ciência aplicada, como é o caso da Linguística Computacional, as técnicas como grafos, árvores, funções etc. são usadas para representar um fenômeno da realidade. Não podemos usar qualquer técnica para representar qualquer fenômeno da realidade. Pode-se representar os significados por grafos porque os significados, conforme se está analisando-os, possuem certas propriedades que os permitem serem representados por grafos. Porém pode haver mais de uma técnica passível de ser usada para representar um fenômeno estudado. A escolha de qual técnica usar deve ser feita tendo em mente as habilidades mais adequadas para nossos fins que essas técnicas possuem.

próxima ao grafo que guarda a propriedade da hierarquia entre seus elementos, não possui algumas habilidades de um grafo úteis para busca orientada a idéia<sup>6</sup>.

O segundo motivo de não hierarquizar os conceitos é que estamos utilizando apenas um único critério que relaciona os conceitos: a especificação. Se, em nosso modelo, um conceito está ligado a outro por uma seta, isso significa que um conceito está especificando o outro. Por um acaso, sabemos pragmaticamente que entre o conceito de um hipônimo e os conceitos de um hiperônimo (como é o caso de “Flamengo” e “time”) há uma relação do tipo *é um*. Essa é uma relação hierárquica. Mas entre outros conceitos a relação pragmática pode não ser hierárquica. Por exemplo, na expressão “gato gordo”, um conceito da estrutura de conceitos de “gordo” terá que se ligar por especificidade (já que a relação argumental especifica os significados das palavras) a um conceito da estrutura de conceitos de “gato”. É muito provável que não possamos admitir que a relação pragmática existente entre esses conceitos seja passível de uma hierarquização. Discutiremos mais esse ponto no capítulo 2, mais exatamente quando se discutir sobre redes semânticas e *frames*.

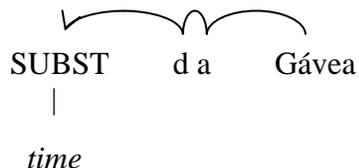
Agora podemos passar para as expressões compostas. A análise semântica das expressões compostas necessita de alguma análise gramatical. Temos as expressões “equipe da Gávea”, “clube da Gávea”, “time da Gávea”, “equipe flamenguista” e “time rubro-negro”, expressando a mesma idéia expressa por “Flamengo”, e as expressões “equipe tricolor”, “time tricolor” e “time das Laranjeiras”, expressando a mesma idéia expressa por “Fluminense”.

Entre “equipe da Gávea”, “clube da Gávea” e “time da Gávea” apenas as palavras “equipe”, “clube” e “time” se alternam, e todas essas palavras detêm o conteúdo significativo representado pelo conceito *time*. Dessa maneira, podemos usar a estrutura de argumentos para definir a estrutura de conceitos. Podemos considerar que a palavra “Gávea” pede um complemento (um substantivo) que possua o conceito

---

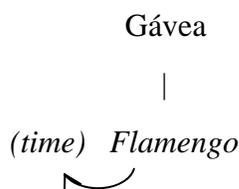
<sup>6</sup> A saber, a habilidade não encontrada na árvore que mais impossibilita o recurso a essa técnica de representação para nossa finalidade é a incapacidade dela de ignorar a seqüência em que seus elementos estão ordenados. Devido à hierarquização de seus elementos, a ordem em que eles aparecem na árvore é fundamental. Isso impediria certas habilidades como a busca de sub-estruturas (como os subgrafos) dentro de estruturas maiores em que se ignorem elementos que aparecem intercalados nessa estrutura ou que se ignore a ordem em que os elementos se dispõem na estrutura.

*time* em sua estrutura de conceitos, sendo que a ligação entre esse complemento e a palavra “Gávea” é mediada pelo artigo “a” e a preposição “de”. Desta maneira:



**Figura 5**

Se o conjunto “equipe/time/clube da Gávea” tem o mesmo significado de “Flamengo”, então esse conjunto gera uma estrutura de conceitos final com o esquema *time Flamengo*. E se é para as palavras “equipe”, “time” e “clube” que é distribuído o conceito *time*, então resta apenas o conceito *Flamengo* para “Gávea”. No entanto, não é apenas esse conceito que forma a estrutura de conceitos de “Gávea”, isso porque o conceito pedido a um complemento e a seta que representa a relação de especificação fazem parte da estrutura de conceitos de uma palavra. Assim, esta será a estrutura de conceitos de Gávea:



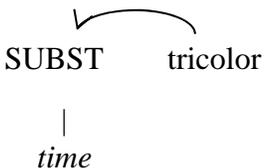
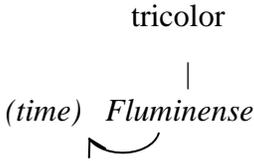
**Figura 6**

O mesmo acontece com “time das Laranjeiras” e seus correlatos.

Estrutura de argumentos	Estrutura de conceitos
<p style="text-align: center;">             SUBST      d as      Laranjeiras                             time           </p>	<p style="text-align: center;">             Laranjeiras                             (time) Fluminense           </p>

**Tabela 1**

No caso de “time rubro-negro” e “equipe tricolor”, o processo é semelhante. A única diferença está na estrutura de argumentos de “rubro-negro” e de “tricolor” que não pedem ligação mediada.

Estrutura de argumentos	Estrutura de conceitos
 <p style="text-align: center;">       SUBST      tricolor                 time     </p>	 <p style="text-align: center;">       tricolor                 (time) Fluminense     </p>

**Tabela 2**

Mas esse é apenas um dos significados de “tricolor” e “rubro-negro”. Há algumas páginas atrás tinha sido apresentada outra estrutura de conceitos para essas palavras, a mesma estrutura de conceitos de “Fluminense” e “Flamengo” respectivamente. São esses sentidos que prevalecem em situações como as das construções “derrota rubro-negra” e “gol tricolor”, pois podem ser parafraseadas por “derrota do Flamengo” e “gol do Fluminense”.

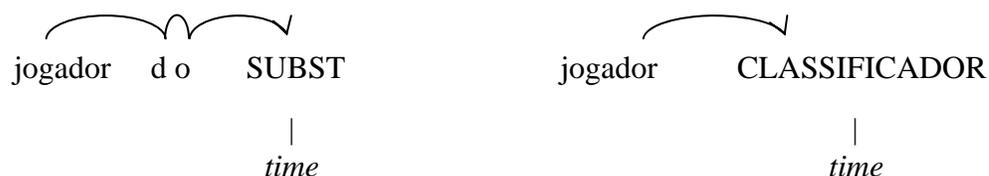
Podemos encerrar por enquanto esse grupo de palavras e passarmos para as palavras que compartilham o sentido de “jogador”. A primeira questão a ser levantada é a respeito da proximidade morfológica entre “jogador”, “jogar”, “jogo”, “jogado” e “jogada”. Essas palavras compartilham estruturas de conceitos?

Vamos tratar por enquanto apenas de “jogador”. No nosso contexto específico, não devemos tratar a relação entre “jogador” e “jogar” como podemos tratar a relação de paráfrases existente entre, por exemplo, “destruir” e “destruidor” das sentenças “José destruiu a casa” e “José foi o destruidor da casa”. No nosso contexto, “jogador” é apenas uma categoria de pessoa, é um tipo específico de pessoa, como o são os técnicos, os árbitros e os torcedores. Isso porque não vamos encontrar uma paráfrase para “jogar”, usando “jogador” em seu lugar.

Então, por ora, não precisaremos nos preocupar com a estrutura de conceitos de palavras como “jogar” para definir a estrutura de conceitos de “jogador”. Mas, como havia sido previsto em outro momento, “jogador” estabelece uma relação com “time”.

Pragmaticamente sabemos que um time é um conjunto de jogadores e que um jogador é um elemento de um time. No entanto, como não foi encontrada nenhuma paráfrase para “time” que contivesse a palavra “jogador” substituindo-a ou substituindo parte de seu significado, então excluimos o significado de “jogador” das palavras que continham o significado de “time”.

A relação entre essas palavras acontece na estrutura argumental da palavra “jogador”, que pede como complemento substantivos que contenham *time*, como ocorre em sentenças como “o jogador do Flamengo”. Esse complemento ainda pode ser um adjetivo-classificador (adjetivo derivado de substantivo), como em “o jogador flamenguista” ou “o jogador rubro-negro”.



**Figura 7**

Com respeito à estrutura de conceitos propriamente dita de “jogador”, podemos seguir passos semelhantes aos seguidos anteriormente e teremos:

PALAVRA	ESTRUTURA DE CONCEITOS	ESTRUTURA DE ARGUMENTOS
jogador	<p style="text-align: center;">jogador   <i>pessoa jogador (time)</i></p>	<p style="text-align: center;">jogador de ART SUBST   <i>time</i></p>
		<p style="text-align: center;">jogador CLASSIFICADOR   <i>time</i></p>
Petkovic	<p style="text-align: center;">Petkovic   <i>pessoa jogador time Fluminense</i></p>	Petkovic
Obina	<p style="text-align: center;">Obina   <i>pessoa jogador time Flamengo</i></p>	Obina
goleiro	<p style="text-align: center;">goleiro   <i>pessoa jogador goleiro (time)</i></p>	<p style="text-align: center;">goleiro de ART SUBST   <i>time</i></p>
		<p style="text-align: center;">goleiro CLASSIFICADOR   <i>time</i></p>
Kleber	<p style="text-align: center;">Kleber   <i>pessoa jogador goleiro time Fluminense</i></p>	Kleber

Tabela 3

Na realidade, quando me referi ao fato de não encontrarmos nenhum exemplo de paráfrases entre “jogador” e “time”, soneguei um caso em particular. Existe um momento em que podemos substituir “jogador” por “time”: as construções em que exista a idéia de “fazer gol”. Por exemplo, sabendo-se que Petkovic é um jogador do Fluminense, podemos parafrasear a frase “Petkovic marcou um gol” por “o

Fluminense marcou um gol”. Mas o contrário não pode ser feito. Não podemos substituir “o Fluminense marcou um gol” por “Petkovic marcou um gol”, nem por “um jogador do Fluminense marcou um gol”, nem mesmo por “um jogador marcou um gol”.

A sentença “o Fluminense marcou um gol” não pode ser parafraseada por “Petkovic marcou um gol” porque não podemos atribuir o gol a Petkovic somente a partir da primeira sentença. O mesmo ocorre com o veto à paráfrase com “um jogador do Fluminense marcou um gol”; não podemos prever que o gol tenha sido feito por um jogador do Fluminense, pois há gols contra. Já o caso da paráfrase com “um jogador marcou um gol” é vetado porque, apesar de a sentença “o Fluminense marcou um gol” pressupor pragmaticamente que algum jogador tenha marcado o gol (afinal somente jogadores podem marcar gols), não podemos parafraseá-las simplesmente porque uma frase como “um jogador marcou um gol” sem qualquer especificação sobre esse gol não ocorre nesse contexto.

Informações óbvias demais não aparecem num jornal. Se uma sentença é escrita para informar sobre um gol, isso vem sempre acompanhado da informação sobre quem o marcou (“gols de Tuta e Diego Souza”), ou sobre que time o marcou (“o Fluminense quase marcou o gol”), ou sobre o modo como o gol foi marcado (“Pet fez um gol de placa”) etc.

Se uma paráfrase não existe num contexto simplesmente porque uma frase potencial nunca ocorre, então não precisamos nos preocupar em prevê-la.

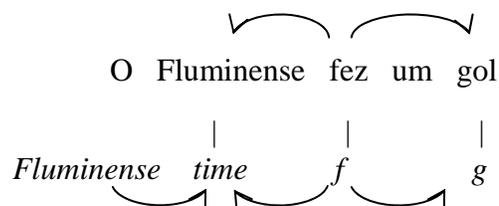
Somente ocorre paráfrase entre uma sentença que informe a qual time pertence o jogador que marcou o gol e uma sentença que informe o nome do time que marcou o gol. Desse modo, a idéia de “Fluminense marcou um gol” está contida em “um jogador do Fluminense marcou um gol” e em “Petkovic marcou um gol”. O termo “Petkovic” possui em sua estrutura de conceitos a informação de que este é um jogador do Fluminense, como mostramos há algumas páginas.

Esse tipo de paráfrase só ocorre com a informação “fazer gol”. Uma frase como “Petkovic se machucou” não pode ser parafraseada por “o Fluminense se machucou”. Essa foi a razão por termos ocultado esse caso de paráfrase entre “jogador” e “time” durante a definição das estruturas de conceitos para as palavras que contivessem essas

idéias. A paráfrase encontrada tem que ser prevista na estrutura de conceitos de palavras que contenham a idéia de “marcar gol”, ficando assim inalteradas as estruturas de conceitos de palavras com a idéia de “jogador” e de “time”.

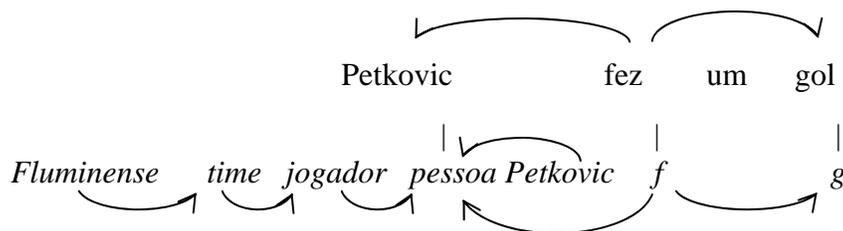
Então começaremos a trabalhar a estrutura de conceitos de palavras ou expressões que contenham a idéia de “marcar gol”. A solução encontrada para a estrutura de conceitos de “marcar gol” incluir a paráfrase encontrada se fez por acréscimos de conceitos e por uma distribuição de ligações entre esses conceitos sem nenhum correspondente a interpretações do significado de “marcar gol”, “jogador” ou “time”. Foi um ato simplesmente mecânico, e por isso vou apresentá-lo de forma mecânica, isso é, com o uso de conceitos representados novamente por letras. Depois as substituiremos por formas mais mnemônicas.

Atribuiremos a cada palavra um conceito e os ligaremos conforme a sintaxe.



**Figura 8**

E o mesmo faremos com a construção com a idéia de “jogador”.



**Figura 9**

Por enquanto, as duas construções não formam paráfrases. Note-se que, da maneira como foram formalizadas, o significado da primeira sentença não está contido na segunda sentença. Para isso acontecer, o grafo que representa a junção das estruturas de conceitos das palavras da primeira sentença deveria ser um subgrafo da

estrutura de conceitos da segunda sentença. Isso não está ocorrendo porque o conceito  $f$  da primeira sentença está ligado a *time* e na segunda sentença está ligado a *pessoa*.

Para resolvermos isso, as estruturas de conceitos de “fez um gol” têm de variar conforme incidam sobre “time” ou sobre “jogador”. Isso significa dizer que a estrutura de conceitos muda conforme a estrutura de argumentos mude.

O primeiro passo para isso é alterar a estrutura argumental de “fazer um gol”. O verbo “fazer” é um verbo que se comporta como “dar” em “dar um abraço”, que na verdade é uma forma analítica de expressar “abraçar”. Em casos como esse, o verbo age como uma palavra de ligação, e não precisa gerar estrutura de conceitos. Veja:

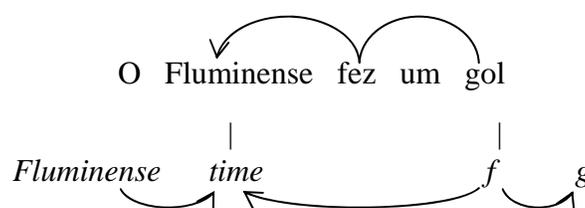


Figura 10

Dessa maneira os conceitos antes atribuídos a “fez” serão agora atribuídos a “gol”. A palavra “gol” também terá de mudar de classe, sendo agora um substantivo-evento. Somente substantivos-evento participam dessa forma de construção. Devemos notar o fato de  $f$  passar a receber a mesma ligação que liga “gol” a seu complemento.

O conceito  $f$  se liga ao conceito *pessoa* de “Petkovic”, impossibilitando a paráfrase. Por esse mesmo processo que é impossibilitado que frases como “Petkovic caiu” e “o Fluminense caiu” não fossem paráfrases, isso é, o as ordenações das relações nos grafos impossibilita que a estrutura de conceitos formado pela sentença “o Fluminense caiu” não contenha a mesma estrutura de conceitos gerada por “Petkovic caiu”.

Por enquanto sabemos que uma das entradas de “gol” terá a estrutura de conceitos e a estrutura de argumentos seguintes:

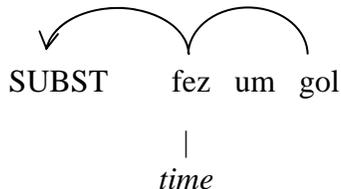
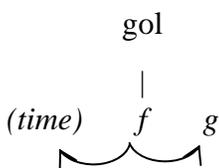
Estrutura de argumentos	Estrutura de conceitos
 <p>SUBST    fez um gol                       time</p>	 <p>gol     (time) f g</p>

Tabela 4

Outra deverá ser a entrada de “gol”, já que pretendemos modificar sua estrutura de conceitos. Lembre-se que diferentes sentidos determinam diferentes entradas. Nessa entrada acrescentaremos um novo conceito *d* em “gol”. Ele se ligará ao conceito *pessoa* do complemento. O conceito *f* continuará na estrutura, ligando-se ao conceito *time* do complemento. Veja:

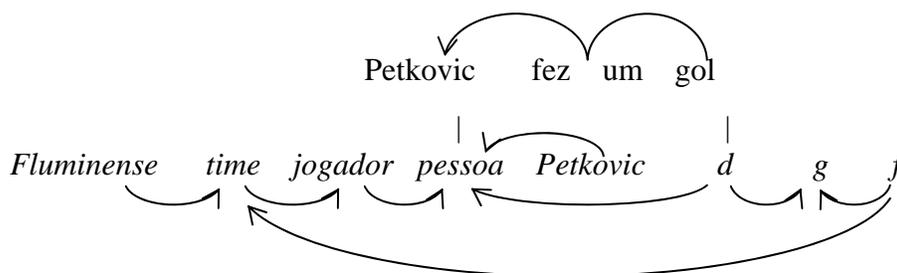


Figura 11

Podemos substituir por formas mais mnemônicas esses conceitos representados por letras. Escolhi substituir *d* por *fazer*, *g* por *gol* e *f* por *possuir*. Essa escolha foi completamente arbitrária, como, aliás, também o foi a escolha de todos os demais nomes de conceitos. As estruturas de conceitos e de argumentos de “gol” ficam assim:

PALAVRA	ESTRUTURA DE CONCEITOS	ESTRUTURA DE ARGUMENTOS
gol		
gol		

Tabela 5

As mesmas estruturas de conceitos são atribuídas a outras palavras que contenham a idéia de “marcar gol” e outras ainda podem ser as estruturas de argumentos de “gol”. Por exemplo, em sentenças como “o gol de Petkovic”.

Passemos agora para a idéia de “jogar”. Novamente devemos saber se em palavras como “jogo”, “jogada” e “jogador” a idéia de “jogar” está presente. Como foi dito antes, “jogador” não contém a idéia de “jogar”, isso porque, se admitíssemos que o significado de “jogar” estivesse contido em “jogador”, toda sentença em que aparecesse a palavra “jogador” conteria a idéia de “jogar”, mesmo em sentenças como “o jogador treinou a manhã inteira” ou, pior, “Petkovic concedeu uma entrevista coletiva”. Mas talvez o oposto seja verdade.

O significado de “jogador” pode estar inserido no significado de jogar, como se “jogar” significasse algo como “ser jogador”. Isso poderia ser usado para permitir paráfrases do tipo que substituísse “Petkovic foi o melhor jogador em campo” por “Petkovic jogou melhor que todos em campo”. Ou talvez esse fosse um caso em que “jogador” possuísse um outro significado, o significado de “jogar”.

É muito comum, no momento em que estamos definindo as estruturas de conceitos das paráfrases, nos debatermos com casos em que mais de uma distribuição é possível, e precisarmos fazer uma escolha. Não existe certo ou errado nas definições de estruturas de conceitos. Existem escolhas que funcionam e outras que não funcionam. E “funcionar” significa tornar o computador capaz de reconhecer

sentenças com as mesmas informações como sendo paráfrases. Isto é, existem escolhas melhores e escolhas piores.

Escolhas que façam com que o computador trabalhe com menos dados serão melhores do que as que exijam mais poder computacional. Portanto, sempre são melhores as distribuições que possuem menos conceitos. Também sempre são melhores as distribuições que geram menos homonímias. E também são melhores as distribuições que conseguem tratar o maior número de paráfrases de uma só vez. O problema é que esses três fatores são conflitantes e muitas vezes conciliá-los é impossível.

Particularmente prefiro tratar o maior número de paráfrases de uma só vez, mesmo que isso gere muitas homonímias e muitos conceitos. Mas isso é uma escolha pessoal, não uma obrigação. Essa alternativa talvez nem seja a mais recomendável para o trabalho em equipe, pois muitas vezes uma representação com muitos conceitos pode ficar bem confusa.

Não precisamos nos preocupar nesse estudo de caso com os eventos em que “jogador” e “jogar” serão paráfrases, já que esses casos não aparecem em nosso corpus. Mas não podemos dispensar as paráfrases entre “jogar” e “jogo”.

Os momentos em que aparecem as palavras “jogo”, “partida”, “clássico” e outras palavras ou expressões semelhantes, elas informam: a) o andamento do jogo (“o jogo melhorou muito”), b) o local onde o jogo ocorreu (“o clássico foi em Volta Redonda”), c) se o jogo terminou (“o fim do jogo”). Em todos esses casos as palavras ou expressões com a idéia de “jogo” devem se relacionar com a idéia de “jogar”. Poderíamos parafrasear os exemplos assim: “jogou-se melhor”, “os times jogaram em Volta Redonda”, “acabou-se de jogar”.

Se quisermos acrescentar outros casos possíveis no uso tratado em que “jogar” e “jogo” formassem paráfrases, poderíamos citar o caso de: “o Flamengo fez um bom jogo” e “o Flamengo jogou bem”. E ainda complicaríamos mais a distribuição de estruturas de conceitos ao parafrasear as construções também por “o Flamengo fez uma boa atuação”.

E ainda existe, nesse corpus, um momento em que “jogo” não compartilha idéias com “jogar”. Isso acontece em casos como “o jogo empatou” e “a partida virou”.

Vamos tratar, portanto, um caso de cada vez. Primeiramente trataremos a estrutura de conceitos de “jogar”, já que ela sempre será parafraseada por “jogo”, enquanto o contrário não ocorre. O primeiro fato a ser notado é que existe uma diferença pragmática entre “time jogou” e “jogador jogou”, no que diz respeito à compreensão que fazemos de “jogar” em cada um dos casos.

Devemos ter cuidado com a compreensão que temos das palavras. Muitas vezes elas nos dão pistas úteis para a composição das estruturas de conceitos, outras vezes, no entanto, nos levam a caminhos enganadores. Nesse caso, a compreensão parece ser uma pista segura, uma vez que ela se relaciona com paráfrases. Em sentenças compostas pela idéia de “o time jogou”, podemos nos referir à atuação do time (“o Flamengo jogou bem”) ou a qual foi o adversário do time (“o Flamengo jogou contra o Fluminense”). Já em sentenças que se componham de “o jogador jogou”, as informações podem informar sobre a atuação do jogador (“Petkovic jogou mal”) ou sobre o fato de o jogador ter sido ou não escalado para o jogo (“Felipe não jogou a partida”). Vamos ignorar somente esse último caso de “jogador jogou” porque não aparece em nosso corpus.

De forma a apresentar o máximo de técnicas de composição de estruturas de conceitos, vamos usar outra estratégia aqui. Usaremos o mesmo conceito para “jogar” em todos os casos apontados. O que vai vetar as paráfrases serão as estruturas de argumento possíveis de “jogar”. Vejamos:

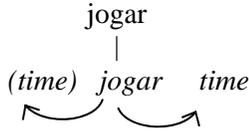
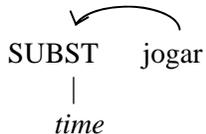
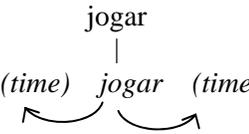
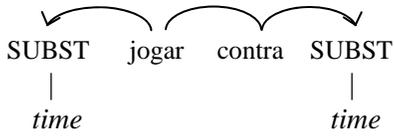
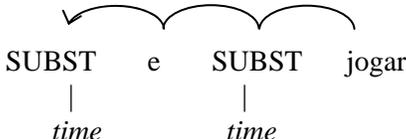
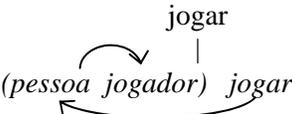
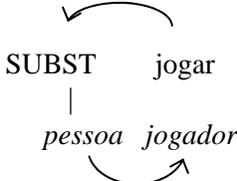
PALAVRA	ESTRUTURA DE CONCEITOS	ESTRUTURA DE ARGUMENTOS
jogar		
jogar		
		
jogar		

Tabela 6

Essas mesmas estruturas de conceitos podem ser repetidas nas palavras sinônimas “jogo” e “partida” que se comportarão como paráfrases em todos os casos de jogar .



Para os outros sentidos de “jogo” e “partida” optamos pelo conceito *partida*.

PALAVRA	ESTRUTURA DE CONCEITOS	ESTRUTURA DE ARGUMENTOS
jogo	jogo   <i>partida</i>	jogo
partida	partida   <i>partida</i>	partida

**Tabela 8**

Esses sentidos de “jogo” e “partida” são usados nos casos “a partida/o jogo virou” ou “a partida/o jogo empatou”. Nesses casos, tanto “virou” quanto “empatou” possuem a idéia de “marcar gol” da forma como já havíamos apresentado.

A idéia de “empatar” é uma das que fazem parte de nosso corpus. Nele, encontramos dois usos para “empatar”. O primeiro uso é caracterizado por “empatar” contendo a idéia de “fazer gol”. Vejam-se os exemplos:

“O empate do Flamengo aconteceu”.

“O time da Gávea conseguiu empatar”.

“Renato empatou a partida”.

“O Flamengo não desistiu e chegou ao empate”.

Nesses casos, poderíamos parafrasear as sentenças com palavras que contivessem a idéia de “marcar gol”.

“O gol do Flamengo aconteceu”.

“O time da Gávea conseguiu marcar”.

“Renato fez um gol”.

“O Flamengo não desistiu e conseguiu o gol”.

É interessante notar que com as substituições não há perda de informações. A informação trazida com “empatar”, nesses casos, traz consigo a idéia “marcar gol” e também pressupõe outro evento “marcar gol”, anterior e atribuído ao time adversário em relação àquele que empatou. Em matérias como as que compõem nosso corpus, nas quais uma partida de futebol é narrada, no momento em que uma sentença como

“Renato empatou a partida” surge no texto, é certo que anteriormente o texto já havia trazido a informação sobre um gol do time adversário. Portanto, se a frase tivesse sido escrita como “Renato fez um gol”, a informação permaneceria inalterada.

No entanto, será que tratar “empatar” e “marcar gols” como paráfrases perfeitas é a melhor escolha para um buscador orientado a idéia?

Isso não seria desejável para o buscador. Imagine que um usuário queira encontrar sentenças que informem que um time fez um gol de empate. Se sempre a informação de “empatar” nesse sentido for decomposta em estrutura de conceitos do mesmo modo como “marcar gol” é decomposta, o buscador retornará para o usuário todas as sentenças em que aparece “marcar gol”, inclusive sentenças como “o Flamengo virou o jogo”. E não era isso o que o usuário esperava.

Podemos pensar na estratégia oposta. Nesse uso lingüístico, sempre que um gol influencia no resultado parcial do jogo, isso é, sempre que um gol puser um time na frente do placar, empatar ou virar a partida, isso será informado explicitamente como tal. Isso quer dizer que, se um gol empata a partida, ele sempre será informado como um gol de empate. A matéria nunca irá informar simplesmente a seqüência de gols, deixando que o leitor vá somando e deduzindo quem está com a vitória, derrota ou empate parciais.

Dessa maneira, não precisaríamos atribuir o significado de “marcar gol” para “empatar” já que a idéia de “empatar” sempre aparecerá expressa com palavras que incluem a idéia de “empatar” e nunca com palavras que somente incluem a idéia de “marcar gol”.

No entanto, acredito que seja uma boa habilidade do buscador encontrar uma sentença como “Petkovic chuta e empata a partida”, quando o usuário requisitar “Fluminense fez gol”. Se assim quisermos, “empatar” deverá conter em sua estrutura de conceitos a idéia de “fazer gol”.

A melhor estratégia então é tratarmos esse problema de forma mecânica. O gol de empate é um tipo específico de gol, que se distingue de um gol que vira a partida ou de um gol que põe o time na frente do placar. Usaremos então apenas um conceito ligado ao conceito *gol* e aproveitaremos todo o resto da estrutura de conceitos de “marcar gol”.

PALAVRA	ESTRUTURA DE CONCEITOS	ESTRUTURA DE ARGUMENTOS
empatar		
empatar		

Tabela 9

Para percebermos a diferença entre esse uso de “empatar” e o outro, será interessante notar também que palavras com a idéia de “vencer” ou “perder” não podem ser usadas nesse caso.

“\*A vitória do Flamengo aconteceu”.

“\*O time da Gávea conseguiu vencer”.

“\*Renato venceu a partida”.

“\*O Flamengo não desistiu e chegou à vitória”.

Note-se que considero como impossíveis as frases acima, considerando-as no mesmo contexto que as originais, isso é significando um momento em que um time ou um jogador fez um gol. Vejamos que, se aceitarmos tais sentenças, a compreensão que fazemos é que elas nos informam sobre um resultado final de um jogo, e, nesse caso, a terceira sentença ainda permaneceria impossível.

A idéia de “resultado final de um jogo” é o segundo uso de “empatar”. Esse caso aparece preferencialmente nas manchetes, no início e no fim das matérias. São momentos em que “vencer” e “perder” poderiam ser usados.

“Clássico Fla-Flu termina empatado”.

“Fluminense e Flamengo empataram em 2 a 2”.

“Em clássico emocionante, Fla-Flu acaba empatado”.

“Flamengo e Fluminense empatam em Volta Redonda”.

“O resultado acabou com um empate emocionante em Volta Redonda”.

Note-se que esses casos não podem incluir a idéia de “marcar gol”, isso porque um empate pode acontecer sem gols. A informação sobre os gols pode ser dada na estrutura argumentativa da palavra que contenha essa idéia de “empatar”. É o que acontece em construções como “empate por 2 a 2”, por exemplo. Esse tipo de estrutura em que aparece o placar é típica desse uso de “empatar” e não ocorre no outro uso.

Nesse uso, a estrutura de conceitos de palavras com a idéia de “empatar” não precisa incluir a estrutura de conceitos de “marcar gol”. Mas é importante notar que a sentença “Flamengo e Fluminense empataram em 2 a 2” deve ser retornada pelo buscador se o usuário requisitar “Fluminense fez gol”. Entretanto, o buscador não deve retornar para o mesmo requisito do usuário sentenças como “Flamengo e Fluminense empataram”, nem mesmo “Flamengo e Fluminense empataram em 0 a 0”. Nesse último caso, nem Flamengo nem Fluminense marcaram gols. Um antônimo ou uma negação de algo nunca podem ser retornados pelo buscador quando esse algo for requisitado pelo usuário.

## **4.2**

### **Problemas ainda não resolvidos pelo modelo**

Como se viu nos exemplos de atribuição de conceitos e relações entre os conceitos apresentados, esse trabalho é bem complicado. A complicação não decorre de ser um trabalho difícil, mas por exigir a atenção para inúmeros detalhes. Quem se propuser a levantar a estrutura de conceitos de uma palavra deve prever todos os seus usos, todas as paráfrases de seus usos, todas as palavras e expressões que podem compartilhar parte de sua estrutura de conceitos em cada uso, todas as palavras e expressões que estão contidas no significado da palavra em todos os seus usos. Trabalhar com tantos dados assim numa situação real de sistemas de recuperação de informações em corpora textuais, cujo número de palavras pode chegar facilmente a dezenas de milhares, é algo que beira a incapacidade humana.

Mas é justamente a capacidade de trabalhar com milhões de dados e variáveis computáveis interagindo entre si que faz da computação a ferramenta útil que é. O

melhor seria encontrar uma forma de automatizar o processo que consiga considerar como paráfrases duas sentenças diferentes. Como já foi frisado inúmeras vezes, a atribuição das estruturas de conceitos é algo mecânico, que muitas vezes é inclusive atrapalhado pelo nosso entendimento das palavras.

O problema que impede a automatização desse processo no estágio atual de desenvolvimento do modelo de representação da semântica para a busca orientada de textos é justamente o fato de ele ainda estar incompleto.

Algumas questões permanecem em aberto, impossibilitando a automatização. Alguns problemas não resolvidos por esse modelo dizem respeito a algumas classes de palavras ainda não trabalhadas, como os advérbios, e também estruturas oracionais, principalmente as orações adverbiais, cujas conjunções devem influenciar as estruturas de conceitos e paráfrases. Por exemplo, deverá haver um conceito para indicar a idéia de causa que se pode notar em alguns eventos, sendo que essa idéia pode ser expressa também em orações adverbiais causativas.

Outros problemas dizem respeito à estrutura textual. O texto também possui uma gramática peculiar. Nos exemplos apresentados de jornalismo esportivo pode-se notar que a manchete tem uma gramática diferente do restante do texto: por exemplo, o artigo é quase que totalmente abolido nas manchetes em momentos em que, se as mesmas frases estivessem escritas no corpo do texto, seria obrigatório o uso do artigo.

A teoria terá que prever essas gramáticas concorrentes.

Em certos momentos, uma desambiguação pode ser feita simplesmente porque uma expressão aparece escrita num determinado espaço do texto. Por exemplo, foram apresentados dois sentidos para a palavra “empatar”, aquele que informava sobre um gol marcado que igualava o placar da partida ou aquele que informava o resultado final da partida. Se o verbo “empatar” estiver escrito na manchete, no início ou no final da matéria, é mais provável que seu sentido seja o de resultado de uma partida. Se o verbo estiver escrito no desenvolvimento da matéria é provável que o sentido seja o outro.

Outro problema latente diz respeito às flexões, principalmente as verbais. Provavelmente, a flexão de tempo terá de ser tratada na estrutura de conceitos,

enquanto a flexão de modo (como talvez também a flexão de número e pessoa) deverá ser tratada na estrutura de argumentos.

Os tempos verbais deverão ser simplificados como ações que ocorreram antes, ao mesmo tempo ou depois de outras ações, não importando se ocorreram no passado, presente ou futuro. Isso porque para o usuário não terá muita importância saber se um evento aparece escrito no presente, no passado ou no futuro. Mas a relação “antes”, “durante” e “depois” entre os eventos é importante para as paráfrases, como é o caso de “empatar”, que poderia ser parafraseada por dois eventos “marcar gol”, sendo que um ocorreria em momento anterior ao outro.

Talvez seja importante informar também se uma ação foi repetida, se prolongou no tempo ou se foi encerrada; essas informações são conseguidas na análise aspectual dos tempos e modos verbais. Portanto temos que teorizar como os diferentes tempos e modos verbais e a interação entre esses tempos e modos de diferentes verbos informam todos esses dados.

Quanto mais paráfrases conseguirmos tratar, mas resultados corretos nosso buscador retornará.

#### **4.2.1 Trabalhos futuros**

O tratamento da semântica entendida enquanto paráfrase se mostra bom para aumentar os resultados relevantes de uma busca bem como restringir os resultados pouco relevantes.

No entanto, como se trata de uma forma de representar a semântica nos arquivos de texto, outros usos podem ser pensados para este modelo. O primeiro deles é a tradução automática. Se pensarmos que uma tradução é uma paráfrase de um texto, por ser uma reescrita em outra língua, entenderemos por que esse modelo pode ser usado para essa finalidade. Para isso, entretanto, não bastaria desenvolver o dicionário eletrônico das duas línguas envolvidas nas traduções, seria necessário desenvolver um dicionário eletrônico da tradução de uma língua para outra. Isso porque as estruturas de conceitos atribuídas às palavras não são universais, são apenas daquela língua e num certo uso específico que se faz dela. Para cada uso se deve

desenvolver um dicionário eletrônico que determine suas paráfrases específicas. A tradução deve ser vista como um uso específico.

Outra utilidade para este modelo é uma espécie de parafraseador automático. Podemos imaginar esse parafraseador como um aperfeiçoamento do recurso que muitos editores de texto possuem de sugerir sinônimos para as palavras destacadas pelo usuário. No parafraseador automático, poderiam ser sugeridas paráfrases para trechos destacados pelo usuário.

Como o tratamento semântico é fundamental para um bom *parser* sintático, esse modelo poderia servir a programas que precisam de uma análise sintática precisa, como parece ser o caso dos revisores gramaticais.

Outro uso do modelo é usá-lo para confeccionar dicionários para rodar em meio digital. Num possível dicionário, o usuário poderia não apenas digitar a palavra cujo significado quer saber ou lembrar ou conferir, como já faz, mas também o usuário poderia ditar o significado cuja palavra correspondente quer saber ou lembrar ou conferir. Ainda poderiam ser feitos dicionários de idéias afins, *thesaurus*, dicionários de sinônimos, ou dicionários de regência a partir de um dicionário eletrônico que seguisse esse modelo.