



Um estudo do *córpus* COMPARA: a semântica dos compostos nominais

Mestranda Lílian Figueiró Teixeira
Bolsista Santander

Dra. Rove Luiza de Oliveira Chishman
Universidade do Vale do Rio dos Sinos
Brasil

Linguatca: 10 anos – Aveiro – setembro, 2008

Compostos Nominais

- Os compostos nominais são extremamente produtivos na língua inglesa.
- Desafio para os sistemas de tradução automática. Principal dificuldade: reconhecer mais de duas palavras como uma unidade.
- Exemplo: *I went to the night school* ⇒ *Eu fui à escola de noite.*

Córpus Paralelo COMPARA

- Fornece dados para o estudo de uma série de fenômenos lingüísticos a partir de equivalências de tradução nas línguas portuguesa e inglesa.
- As linhas de concordância do COMPARA servem para o estudo da semântica e para a identificação de padrões de tradução dos compostos.

Objetivos

- Estudar a semântica dos compostos nominais formados por dois substantivos (NN), tendo como ponto de partida a sua tradução do inglês para o português.
- Identificar padrões semânticos de tradução destes compostos.
- Optar por uma tipologia semântica que dê conta dos compostos encontrados no **cópus**.

Etapas

- Extração de compostos em inglês freqüentes no COMPARA.
- Obtenção das linhas de concordância do sítio.
- Identificação dos equivalentes de tradução.
- Análise semântica dos compostos.
- Sistematização dos resultados.

Extração dos compostos

- Um composto formado por dois substantivos (NN) em inglês apresenta um **pré-modificador** seguido por um substantivo **núcleo**.
- Tanto o modificador quanto o núcleo podem ser o **nódulo** (*core word*), já que é esta palavra que vai ser encontrada em outros compostos, formando *família de compostos*.

Extração dos compostos

- Busca por linhas de concordância em que dois substantivos aparecem juntos.
- Consideramos tanto os substantivos no singular quanto no plural.
- Através da fórmula **[pos="N.*"][pos="N.*"]** digitada na busca avançada, obtivemos **32.216 ocorrências**.
- Os seguintes nódulos recorrentes foram selecionados: *hall, room, house, door, floor, table, window e school*.

Linhas de concordância

- Busca através da fórmula `[pos="N.*" & word="school"] @[pos="N.*"]`.
- Os diferentes nódulos são digitados onde se encontra a palavra *school*.
- A ordem do nódulo também foi invertida e as expressões com um número maior de resultados foram selecionadas.
- Os compostos formados por mais de dois substantivos e os com algum elemento deverbal (como *-ing*) foram excluídos.

Dados do COMPARA

Composto	Ocorrências	Exemplo
N hall	67	<i>concert hall</i>
N room	78	<i>hotel room</i>
N house	226	<i>country house</i>
N door	135	<i>kitchen door</i>
N floor	54	<i>ground floor</i>
N table	94	<i>dinner table</i>
N window	55	<i>train window</i>
school N	47	<i>school gate</i>
N school	59	<i>summer school</i>

Equivalentes de tradução

- *guest room* – 5 ocorrências: quarto de hóspedes (3), quarto do hóspede (1) e quarto para hóspedes (1);
- *country house* – 12 ocorrências: casa de campo (2) e quinta (10);
- *glass door* – 6 ocorrências: porta envidraçada (2), porta de vidro (2), portinha envidraçada (1) e SEM TRADUÇÃO (1).

Análise dos Dados

Composto	Exemplos	Tradução	Relação semântica
N hall	concert hall school hall	sala de concertos salão da escola/refeitório	papel télico papel constitutivo
N room	hotel room laundry room	quarto de hotel quarto de engomados/lavanderia	papel constitutivo papel télico
N house	station house summer house family house brick house beach house hen house	delegacia casa de verão casa da família casa de tijolo casa da praia galinheiro	papel télico tempo posse papel constitutivo local papel constitutivo
N door	trap door glass door street door garden door	alçapão porta de vidro/porta envidraçada porta da rua porta que dava para o jardim	papel constitutivo papel constitutivo local local

Análise dos Dados

Composto	Exemplos	Tradução	Relação semântica
N floor	ground floor kitchen floor metal floor	andar térreo chão da cozinha chão metálico	local papel constitutivo papel constitutivo
N table	kitchen table bedside table dinner table coffee table tin table	mesa da cozinha mesa-de-cabeceira mesa de jantar mesinha mesa metálica	papel constitutivo local papel télico papel télico papel constitutivo
N window	kitchen window picture window ticket window	janela da cozinha janela panorâmica guichê	papel constitutivo papel télico papel télico
school N	school holiday school gate school report	férias portão do colégio boletim escolar	tempo papel constitutivo local
N school	summer school night school Sunday school	curso de verão escola noturna escola dominical/catequese	tempo tempo tempo

Análise dos Dados

- A relação entre os dois substantivos de um composto, na maioria dos casos, pode ser explicada através de dois papéis da estrutura *qualia* (Pustjovsky, 1995), o constitutivo e o télico.
- Télico: *dinner table*.
- Constitutivo: *school gate*.

Análise dos Dados

- Grande parte das traduções dos compostos para o português é N de N. No entanto, a preposição “de” possui diversos significados (além dos papéis constitutivo e télico):
 - posse - *family house*
 - tempo - *summer school, sunday school*
 - local - *street door*

Tradução N Adjetivo

- Se é possível traduzir o composto de duas formas, N de N ou N Adjetivo, os dois casos são encontrados no *córpus*.
- Geralmente o uso do adjetivo está relacionado a algum material do qual o objeto é feito.
- Em *metal floor* e *tin table*, os equivalentes de tradução são *chão metálico* e *mesa metálica*.
- Quando não há um adjetivo correspondente em português para o material, mantém-se a construção N de N (*brick house*).

Tradução N

- Quando existe uma única palavra em português correspondente ao composto em inglês, o seu uso é preferido.
- Enquanto há três ocorrências para *lavanderia*, *quarto de engomados* só aparece uma única vez.
- Outros equivalentes são escolhidos, pois se percebe certo grau de lexicalização no seu uso. Em *Coffee table*, o seu equivalente é *mesinha*. Mesmo sendo um composto télico, a informação mais importante aqui não é o seu uso, mas o seu tamanho.

Comentários

- O estudo aqui empreendido e a definição de uma tipologia semântica para descrever os compostos nominais do tipo NN em inglês e seus correspondentes em português podem servir de base para pesquisas voltadas para o aprimoramento de sistemas de tradução automática.
- O acesso a um corpus paralelo se mostrou útil para um estudo bilíngüe, podendo contribuir para outros estudos sobre diferentes fenômenos lingüísticos.
- Sugestões para um futuro estudo: analisar como os compostos são traduzidos do português para o inglês e observar quais os equivalentes de tradução em inglês dos compostos formados por NdeN na língua portuguesa.

Um estudo do **córpus** **COMPARA: a semântica dos** **compostos nominais**

Lílian Figueiró Teixeira

lilianjoy@gmail.com

Dra. Rove Luiza de Oliveira Chishman

rove@unisinis.br

Universidade do Vale do Rio dos Sinos
- Brasil



Linguatca: 10 anos – Aveiro – setembro, 2008
