The annotation of causative constructions in a dialectal corpus

A mãe fez comer a sopa.

Faire

The mother made the kids eat the soup.

1. INTRODUCTION

- Syntax-Oriented Corpus of Portuguese Dialects http://www.clul.ul.pt/presources41-cordial-corpus

- annotated corpus: (almost) spontaneous speech

- 42 locations / 600 000 words (~ 68 hours of audio recording)

- Informants: aged, non-educated, rural, born and raised in the place of the interview

- Syntactic annotation (under development: ~200.000 words)

- Automatic search (CorpusSearch2: http://corpussearch.sourceforge.net)


- CF. Carrilho 2010; Magro 2010

2. CORDIAL-SIN Annotation system

(Tycho Brahe corpus; Penn parsed corpus)

- Input: POS annotated texts

- A/F mãe/N fez/VB-D comer/VB a/D-F sopas/N a@IP @os/D-P filhos/N-P @P.

- Causative Constructions

- Inflected vs non-inflected infinitive

- position and form of the causative:
  - Null subject "pro" or NP between the verbs
  - (Faz os porcos/"pro" ofar a carne) (CLH23)

- "It makes the pigs get fat"

- Null subject "arfb":

- (Deixa esta vara crescer, (PST01)

- "Let this twig grow"

- (Deixe ir a sua Ester, (GRU31)

- "Let your Ester go"

- PP subject between and after the verbs:

- (Mandaram-lhe arrumar um caminetão ali) (ALL34)

- "We sent them to grind"

- Empty category coined with a clinic:

- (Mandaram-lhe arrumar um caminetão ali) (ALL34)

- (Deixa ir a sua Ester, (GRU31)

- (Mandaram-lhe arrumar um caminetão ali) (ALL34)

- "They sent them to park the van there"

- (Ela mandou-o lá ir, (COV11)

- "She sent him there to go there"

3. Methods

- Automatic search (CorpusSearch2: http://corpussearch.sourceforge.net)

- CorpusSearch2 (http://www.clul.ul.pt/resources/411-cordial-corpus)

- 42 locations / 600 000 words

- POS tags and a rich syntactic annotation scheme allow

- almost) spontaneous speech

- The annotation of causative constructions in CORDIAL-SIN”, presented at First Workshop on Syntactic Variation, IKER, Bayonne, 2010

- Piparalho, Catarina. 2010. "When CORDIAL becomes a friendly: endowing the CORDIAL-SIN corpus with a syntactic annotation layer

- First Workshop on Syntactic Variation, IKER, Bayonne, 2010

Contactos
Av. Professor Gama Pinto, 2, 1649-003 Lisboa, Portugal
Tel.: +351 217 904 700 | Fax: +351 217 965 822 | www.clul.ul.pt