Learning to Assign Grammatical Function Labels

Grzegorz Chrupała   Josef van Genabith

National Center for Language Technology
School of Computing
Dublin City University

NCLT Seminar Series, 15 Feb 2006

Outline

1. Motivation
2. Learning Function Labels
3. Evaluation

Outline

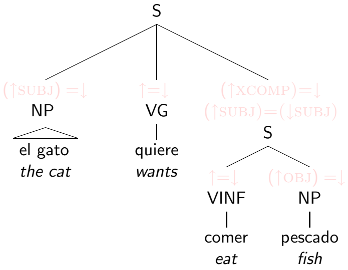
GramLab project

- Treebank-based Multilingual Lexical Functional Grammar induction for English, Chinese, Japanese, Arabic, German, French and **Spanish**
- Leverage successful data-driven approaches to parsing
- Core idea: automatic annotation of constituent trees (c-structures) to obtain LFG f-structures

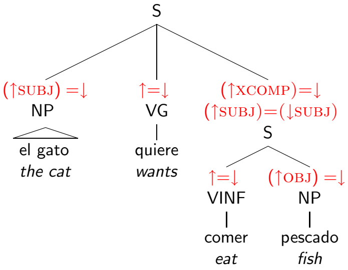
What are f-structures

- F-structures encode grammatical relations such as Subject, Object, Sentential Complement etc.
- Less interlanguage variability than c-structure
- Explicit representation of
  - control and raising constructions
  - pro-drop
  - long distance dependencies
- Syntactic interface to predicate structure and semantics

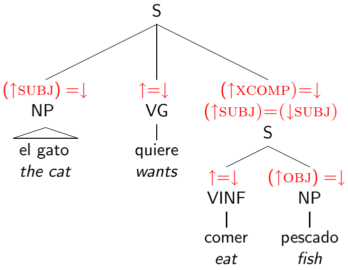
C-structure to f-structure mapping



C-structure to f-structure mapping



C-structure to f-structure mapping



$$\left[\begin{array}{ll} \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`gato'}\end{array}\right] \\ \text{PRED} & \text{`querer}\langle\text{SUBJ,XCOMP}\rangle\text{'} \\ \text{XCOMP} & \left[\begin{array}{ll} \text{SUBJ} & [\ \ ] \\ \text{PRED} & \text{`comer}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{`pescado'}\end{array}\right] \end{array}\right] \end{array}\right]$$

Automatic functional annotation

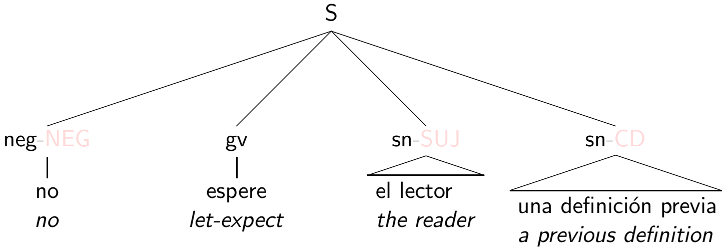
Initial work done on English Penn II Treebank

- Extract most common CFG rules from treebank
- Write annotation rules based on node context:
  - Node category
  - Node's mother's category
  - Position relative to head: head / left / right
- Apply to treebank trees or to parser output
- Collect and solve functional equations

Issues with Spanish

Difficulties in applying this approach to Spanish and Cast3LB

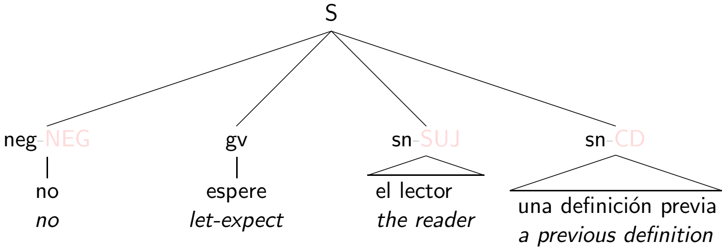
- flexible order of sentence constituents
- information encoded in terms of tree configurations in Penn, but in function tags in Cast3LB



Issues with Spanish

Difficulties in applying this approach to Spanish and Cast3LB

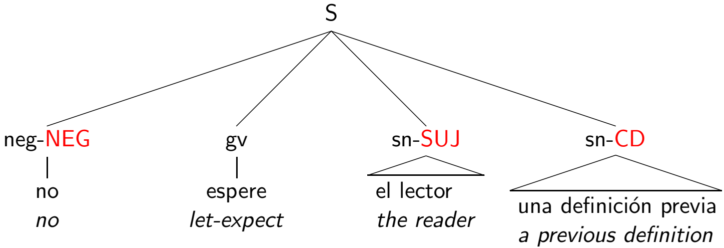
- flexible order of sentence constituents
- information encoded in terms of tree configurations in Penn, but in function tags in Cast3LB



Issues with Spanish

Difficulties in applying this approach to Spanish and Cast3LB

- flexible order of sentence constituents
- information encoded in terms of tree configurations in Penn, but in function tags in Cast3LB



Arguments vs adjuncts

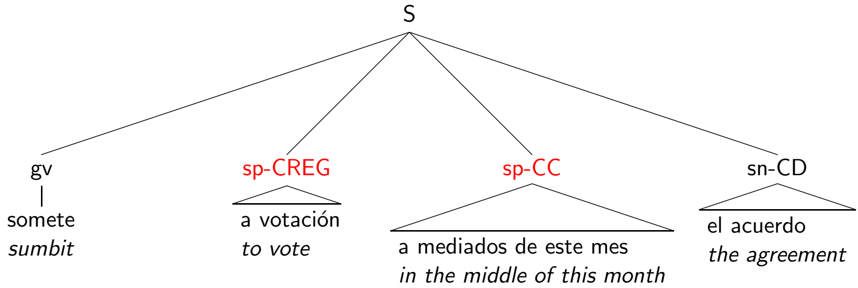
A general problem: argument-adjunct distinction

1. can be difficult to make
2. sometimes is arbitrary

Arguments vs adjuncts

A general problem: argument-adjunct distinction

1. can be difficult to make
2. sometimes is arbitrary



Previous work on automatic functional annotation for Cast3LB

- O'Donovan et al. 2005 decided to rely on Cast3LB functional tags.
- Bikel's parser trained to output function tags by treating compound labels as atomic
- Reasonable results, but maybe we can do better.

Outline

Adding function tags to parse trees

- Can better results be obtained if function tags are added in a postprocess? Advantages:
  - Smaller category label set resulting in reduction of data sparseness
  - Control over learning algorithm and feature set used
- Previous work, Penn Treebank, Charniak's and Collins's parsers
  - Blaheta and Charniak, Assigning function tags to parsed text, *NAACL 2000*
  - Jijkoun and de Rijke, Enriching the output of a parser using memory-based learning, *ACL 2004*

Machine-learning approach

- Adding function tags as a classification task
  - Classes: corresponding to the 24 Cast3LB tags plus the null class for untagged nodes
  - Candidate nodes: sisters to nodes labelled **gv** (Verb Group), **infinitiu** (Infinitive) and **gerundi** (Gerund)
- Experiments with two algorithms: memory-based (Timbl) and Support Vector Machines (R interface to LIBSVM)

Node Features

Nine features: configurational, morphological and lexical information about candidate node

- Position relative to head
- Node head lemma
- Alternative head lemma
- Node head POS
- Node label
- Is definite NP
- Is NP and agrees with head verb
- No. of terminal descendents
- Head lemma +human

Local and Context Features

Four **local** features (head verb and parent information)

- Head verb lemma,
- Head verb person,
- Head verb number,
- Parent node label

8×4 **context** features (= node features of two previous and two following nodes in linear order)

For SVM, feature values of frequency < 20 were replaced with a dummy value

Outline

Evaluation metrics

- Accuracy: proportion of **candidate** nodes that were assigned correct label
- Precision, Recall and F-score defined on sets of tuples: $\langle GF, i, j \rangle$, where
  - GF – grammatical function
  - $\langle w_i, \ldots, w_j \rangle$ – subsequence of tokens spanned by **function-tagged node**

Example:

$[_{NEG}$ No $]$ espere $[_{SUJ}$ el lector $]$ $[_{CD}$ una definición previa $]$

no let-expect the reader a previous definition

$\{\langle NEG, 1, 1 \rangle, \langle SUJ, 3, 4 \rangle, \langle CD, 5, 7 \rangle\}$

Evaluation metrics

- Accuracy: proportion of **candidate** nodes that were assigned correct label
- Precision, Recall and F-score defined on sets of tuples: $\langle GF, i, j \rangle$, where
  - GF – grammatical function
  - $\langle w_i, \ldots, w_j \rangle$ – subsequence of tokens spanned by **function-tagged node**

  Example:

  $[_{NEG}$ No] espere $[_{SUJ}$ el lector] $[_{CD}$ una definición previa]

  no *let-expect* *the reader* *a previous definition*

  $\{\langle NEG, 1, 1\rangle, \langle SUJ, 3, 4\rangle, \langle CD, 5, 7\rangle\}$

Results for gold-standard trees

	Accuracy	Precision	Recall	F-score
Timbl	87.55	88.06	87.66	87.86
SVM	**87.64**	88.15	87.75	**87.95**

Results for parser output

Parser performance

	Labelled Bracket P	Labelled Bracket R	F
All	79.52	79.10	79.30
< 40	82.09	81.49	81.78

Labelling

	Precision	Recall	F-score
Baseline	51.97	51.68	51.82
Timbl	56.54	56.57	56.55
SVM	56.91	56.78	**56.84**

Baseline – parser treating compound category-function labels as atomic.

Results for parser output

Parser performance

	Labelled Bracket P	Labelled Bracket R	F
All	79.52	79.10	79.30
< 40	82.09	81.49	81.78

Labelling

	Precision	Recall	F-score
Baseline	51.97	51.68	51.82
Timbl	56.54	56.57	56.55
SVM	56.91	56.78	**56.84**

Baseline – parser treating compound category-function labels as atomic.

Parse quality influence

- Features used refer not only to candidate node but to mother and sister nodes
- Small constituency misconfiguration may influence classification decision of several nodes
- But, small improvements in parse quality may give greater improvements in function tag assigment :-)
- In a recent experiment:
  - Evalb f-score went up to **80.68 (+1.38)**.
  - Labelling f-score went up to **59.6 (+2.76)**.

F-Structure evaluation

Cast3LB tags to LFG functions is not a one-to-one mapping, eg.:

- CD $\rightarrow$ {OBJ, COMP, XCOMP}
- {CC, MOD, ET} $\rightarrow$ ADJUNCT

F-structure evaluation less sensitive to slightly incorrect parse trees.

Results for parser output with SVM tags – 100 gold standard f-structure-derived dependencies

Precision	Recall	F-score
76.6	75.58	76.09

Error analysis

Simplified confusion matrix for SVM predictions on gold trees

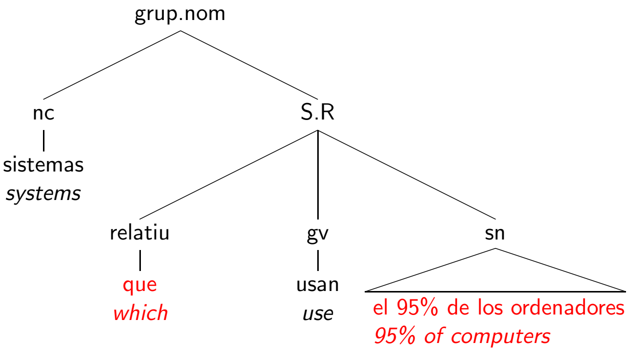
	SUJ	**CD**	**CI**	**CC**	**CREG**	**ATR**
SUJ	464	29	1	10	0	0
CD	29	410	5	22	4	1
CI	0	1	44	8	0	0
CC	5	15	12	539	29	5
CREG	0	0	0	8	39	0
ATR	5	0	0	3	0	137

Subject vs Direct Object

- Over 55% of mistakenly tagging CD as SUJ are daughters of relative clause nodes (S.R), against 16% of all candidate nodes.
- Subject - Direct Object ambiguity in relative clauses

Subject vs Direct Object

- Over 55% of mistakenly tagging CD as SUJ are daughters of relative clause nodes (S.R), against 16% of all candidate nodes.
- Subject - Direct Object ambiguity in relative clauses



Adjunct vs Prepositional Object

In addition to cases of ambiguity

- Training set too small to learn many lexical dependencies
- Inherently fuzzy distinction
- Subtle semantic distinctions and/or a degree of arbitrariness
- Compare
  - Llegar [$_{CREG}$ a un acuerdo]
    arrive *at an agreement*
  - Llegar [$_{CC}$ a Estados Unidos]
    arrive *in the US*

Further work

- Unbounded dependencies between argument and verb
- Inter-annotator agreement for function would give us an idea of where the upper bound is.
- What to do about arbitary distinctions and inconsistencies in human annotation?