Modeling Attachment Decisions with a Probabilistic Parser

Ulrike Baldewein[1], Frank Keller[2]
[1]Saarland University, [2]University of Edinburgh
ulrike@coli.uni-sb.de

Introduction

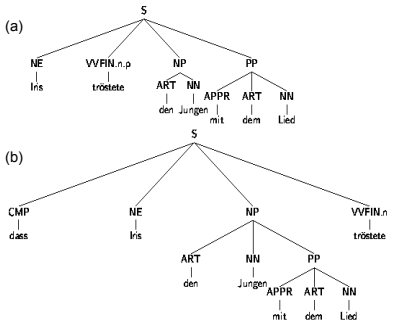
Probabilistic parsing models have been used successfully to model attachment decisions in sentence processing (e.g., Jurafsky 1996, Crocker and Brants 2000, Sturt et al. 2003).
However
- Most models focus on a small selection of phenomena
- All existing models deal exclusively with English data.

- We describe a **broad coverage probabilistic model** of human parsing for German.
- We also make incremental predictions for the attachment decisions for **PP attachment ambiguities**.
- We look at a new phenomenon in PP attachment: **German verb final clauses**

German PP Attachment

- PPs can be attached to 1(a) the verb or 1(b) the preceding NP
- Attachment is influenced by verb subcategorization preference and semantic plausibility

Fig. 1: PP Attachment



German has two basic word order variations for sentences:

- Verb second sentences like 1(a):
  - "Iris comforted the boy with the song"
  - Word order is identical to English
  - The PP is preferentially attached according to the verb's subcategorization frame, as in English (Konieczny et al. 1997)
  - Here: Attachment to the verb, as it subcateogorizes for an NP and a PP (VVFIN.n.p)
  - Attachment is semantically plausible

- Verb final sentences like 1(b):
  - "that Iris comforted the boy with the song"
  - Word order: The verb ("tröstete") comes last, while the order of the other constituents remains the same.
  - The PP is processed before the sentence head is read
  - Preferential initial attachment is to the (seen) NP (Konieczny et al. 1997)
  - Here: Initial attachment preference is to the NP (later supported by verb preference: VVFIN.n)
  - But: NP attachment is not semantically plausible; Verb attachment is forced

The Model

Our model consists of two fully probabilistic modules.

- Syntactic module:
  - Statistical parser which guarantees broad coverage (98% on unseen data)
  - Uses verb subcategorization preferences and general attachment preferences (phrase rule preferences)
  - Ranks alternative structures when the PP has been processed; higher ranked structure is preferred

- Semantic module:
  - Emulates semantic disambiguation of attachment; more plausible attachment is preferred
  - Uses standard selectional preference measure (Clark and Weir 2002) and co-occurrence metrics (Volk 2001) from computational linguistics to evaluate the PP attachment
  - Selectional preference measure evaluates the fit of the PP's head noun as a PP argument of the verb
  - Co-occurrence metric decides for the more frequent combination of verb, preposition and head noun
  - Decision rule:
    - Compare output of selectional preference measure for verb - PP head noun to the attachment threshold as determined on the development set (see next section)
    - If value is below threshold, stipulate NP attachment, else verb attachment
    - If the selectional preference method does not return a result, back off to co-occurrence metric:
    - If $\frac{freq(NPnoun, prep, PPnoun)}{freq(NPnoun)}$ is greater than $\frac{freq(verb, prep, PPnoun)}{freq(verb)}$, choose NP attachment; otherwise, choose verb attachment

- Conflicts between the decisions of the two modules (differences in rank) are interpreted as conflicts between verb subcategorization preference and semantic plausibility. They predict longer reading times.

- The model was evaluated against average reading times on the PP from Konieczny et al. (1997) (no significant effects were reported on the verb).

Training and Test Data

- The syntactic module was trained on 18,000 of the 20,000 sentences in the NEGRA corpus, a syntactically annotated corpus of German. The remaining 2,000 sentences were used for parameter tuning and testing.
  - Pre-tests showed that the **NEGRA corpus has a preference for PP attachment to the verb**

- The semantic module requires a far greater amount of training data
  - The Frankfurter Rundschau corpus was used for the Clark and Weir (2002) method
  - The WWW was used for the Volk (2001) metric

- The items from Experiment 1 and 2 of the Konieczny et al. (1997) paper were split into
  - Development set: Used to set the attachment threshold for the selectional preference measure.
  - Previously unseen test section: Used to evaluate the semantic module

Results

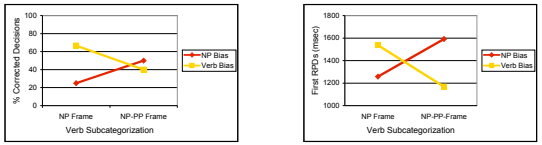


Fig. 2: Exp.1, verb second: Predictions of our model (left) in comparison to the Konieczny et al. data (right)

Fig. 2 shows that the model **correctly accounts for** attachment preferences in **verb second sentences**: When the preferred attachment alternative is semantically implausible, the model predicts longer reading times, which is indeed the case in the experimental data. This replicates modeling results for English (Jurafsky 1996, Crocker and Brants 2000, Sturt et al. 2003).
However, the model **fails to correctly account for** the attachment preferences in **verb final sentences** (Fig.3). It predicts parsing difficulty if the (initially preferred) NP attachment proves plausible, while the experimental data show these cases to be simple.

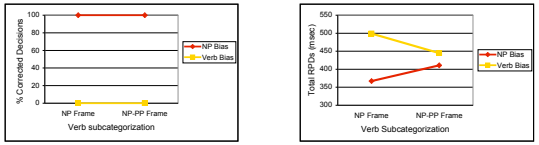


Fig. 3: Exp1, verb final: Predictions of our model (left) in comparison to the Konieczny et al. data (right)

The model's performance for the items from Experiment 2 mirrors these results (Baldewein 2003).

Discussion

To resolve the attachment, the syntactic module relies on verb subcategorization information and on a global attachment preference if the verb is absent. The **experimental data** from Konieczny et. al (1997) show a **global preference for NP attachment**, while in our corpus, verb attachment is more frequent.
Therefore, the syntactic module consistently makes a wrong prediction for verb final sentences, which compromises the model's performance.
However, our model would have been successful had the global preference in the corpus data been for NP attachment. This indicates that the new phenomenon of **PP attachment in verb final sentences can in principle be covered** by the probabilistic framework.
Our results highlight how sensitive probabilistic models are to idiosyncrasies in the training data. Note that in general, balanced corpora consisting of data from different sources are more reliable than newspaper-only corpora like NEGRA.

References:

Baldewein, U. (2003). Modelling Attachment Decisions with a Statistical Parser. Master's Thesis, University of Edinburgh
Clark, S. and Weir, D. (2002). Class-based probability estimation using a semantic hierarchy. Computational Linguistics, 28(2), 187-206.
Crocker, M. and Brants, T. (2000). Wide-coverage probabilistic sentence processing. Journal of Psycholinguistic Research 29(6), 647-669.
Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. Cognitive Science 20,137-194.
Konieczny, L., Hemforth, B., Scheepers, Ch. and Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. Language and Cognitive Processes 12(2/3), 307-348.
Sturt, P., Costa F., Lombardo, V. and Frasconi, P. (2003). Learning first-pass structural attachment preferences with dynamic grammars and recursive neural nets. Cognition 88 (2), 133-169.
Volk, M. (2001). Exploiting the WWW as a corpus to resolve PP attachment ambiguities. In Proceedings of Corpus Linguistics 2001, Lancaster.