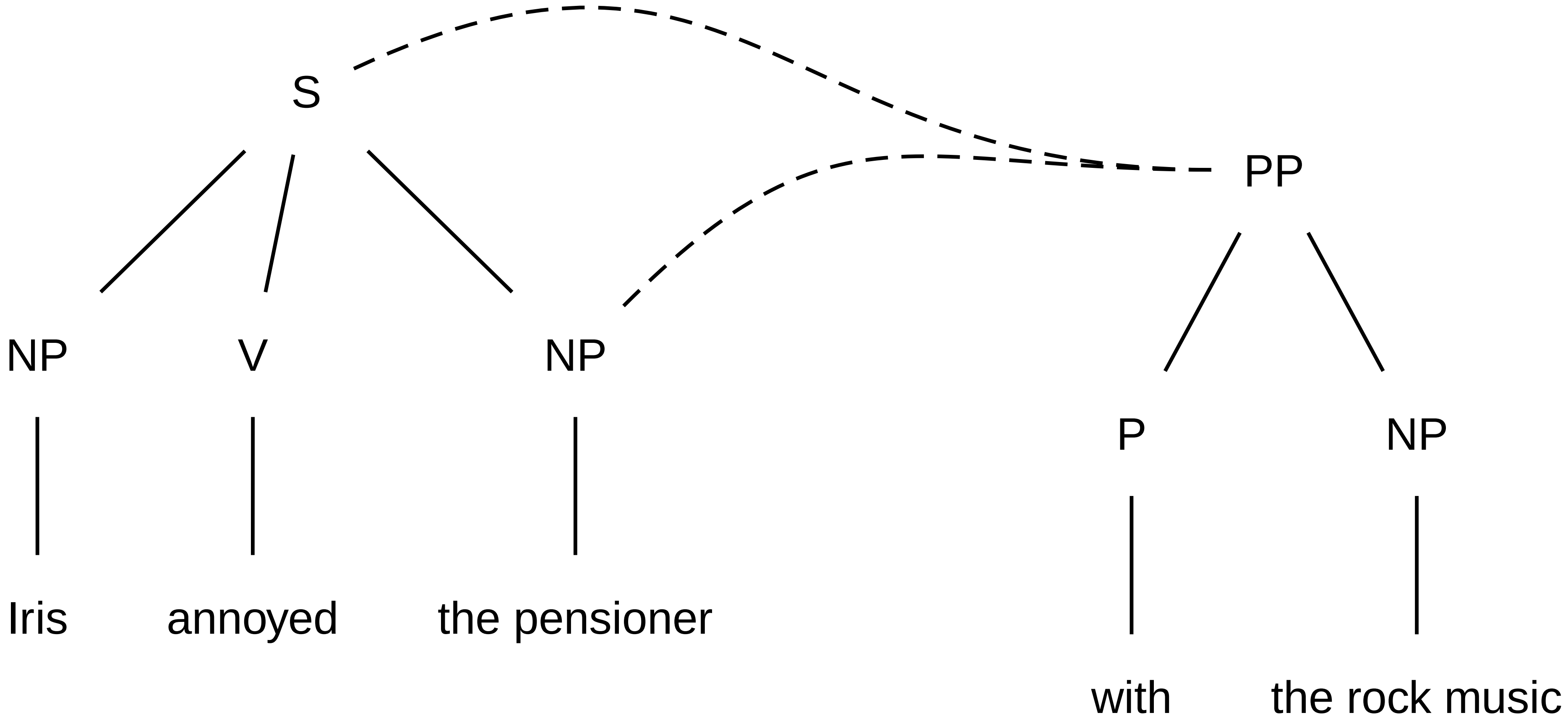
(Beginning to) Model Semantic Aspects of Sentence Processing

Ulrike Baldewein

IGK Colloquium

Semantics in Sentence Processing



Motivation

- Standard models of human reading behaviour cover syntactic phenomena:
  - NP/S ambiguity *The athlete realised her goals were out of reach*
  - MC/RR ambiguity *The horse raced past the barn fell*
  - Lexical ambiguity *The old man the boats*
- How do we model semantic aspects?

  The horse raced past the barn fell is hard and

  The horse led past the barn fell is not

Talk Overview

- Master's Thesis:

  Modelled syntactic and semantic aspects of PP attachment ambiguity

- PhD outline:

  Plan to model initial semantic processing through semantic role assignment:

  More general, more principled approach to the modelling of semantic influence on processing

MSc Thesis – Overview

- PP attachment ambiguity
- Human Behaviour
- The Model
- Final Results
- Conclusions

Attachment Ambiguity

- PPs can be legally attached to either NPs or VPs in German and English
- *Iris annoyed [the pensioner with the rock music].* vs
  Iris annnoyed [the pensioner] [with the rock music].
- This causes ambiguity - so how do humans decide the attachment?

Eyetracking Study

- Konieczny et al. 1997 did an eyetracking study of PP attachment in German
- Tested verb second and verb final sentences
    - *Iris störte den Rentner mit der Rockmusik*
    - *. . . , daß Iris den Rentner mit der Rockmusik störte*
- Varied verb subcategorisation
- Attachment was disambiguated by semantics

Eyetracking Results

- Verb second sentences: When verb subcategorisation and semantic bias clash, reading times increase

  $\Rightarrow$ Initial Attachment is influenced by verb subcategorisation

- Verb final sentences: When semantic bias is for verb attachment, reading times increase

  $\Rightarrow$ Preferential initial attachment is to the NP

The Model – Overview

- Probabilistic model of sentence processing (symbolic backbone)
- Modular
- Unsupervised: Modules are trained separately, then combined (no direct training of the full model on the experimental items)
- Broad coverage

The Model – Architecture

- Two Modules: Syntactic and Semantic
- Both modules make attachment predictions when the PP is encountered
- Syntactic module uses verb subcategorisation and parse tree probability
- Semantic module uses thematic fit of verb and PP or co-occurence patterns
- If the modules differ, we predict longer reading times in humans

Training and Evaluation

- Train the syntactic module on a corpus (NEGRA)
- Train the semantic module methods on a bigger corpus (FR corpus/WWW)
- Split the Konieczny et al. experimental items into development and test set
- Compare the semantic module methods and determine thresholds on the development set
- Evaluate the model on the test set

Syntactic module

- Models syntactic preferences
- Statistical parser
  - Grammar and lexicon read off the 20,000 sentence NEGRA corpus
  - Grammar includes verb subcategorisation information
- Gives broad coverage by being able to process unseen text accurately

Syntactic Module: Evaluation

- Coverage: 98% of unseen test sentences from the NEGRA corpus can be parsed
- Precision: 66.7, Recall: 63.9

  (Dubey & Keller 2003: P=71.3, R=70.9)
- Attachment prediction:
  - Baseline: 50% (Half the items show verb attachment, half NP attachment)
  - Verb second: 42.8% correct
  - Verb final: 50% correct

    But: got the wrong 50% right! Always predicted attachment to the verb

Semantic Module

- Model semantic fit of the attachment through selectional preferences of the verb (Clark & Weir 2002)
- Use co-occurrence measure (Volk 2001) as a backoff if
  - Selectional preference method is not applicable (no verb seen)
  - Selectional preference method does not return a result

Semantic Module – Selectional Preferences

- Clark & Weir method traverses an ontology to find the ideal class for the argument head given the verb and returns an association measure
- Ideal class avoids sparse data problems in computation but does not overgeneralise
- Counts for the model are derived from FR corpus, ontology is GermaNet
- Decision for or against attachment depends on an attachment threshold

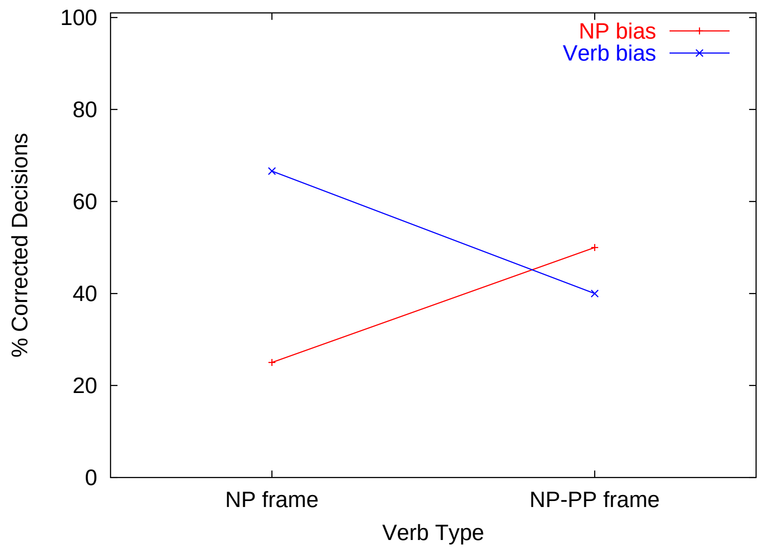
Semantic Module – Co-occurence

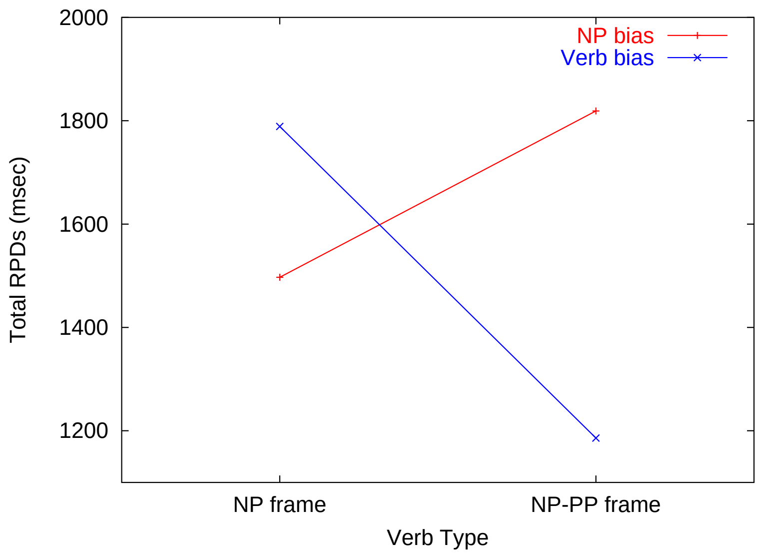
- Volk method counts the times the verb (or NP) and PP head have been seen together in a PP
- Huge sparse data problem: Use the WWW
- Approximate syntactic structure by string query
- Attach towards the site (verb or NP) with the higher co-occurrence count

Semantic Module – Evaluation

- Baseline: 50%
- Clark & Weir method alone: 70% correct attachments (where applicable, 50% coverage)
- Volk method alone: 64.3% correct attachments (100% coverage)
- Combination: 66.6% correct attachments (100% coverage)

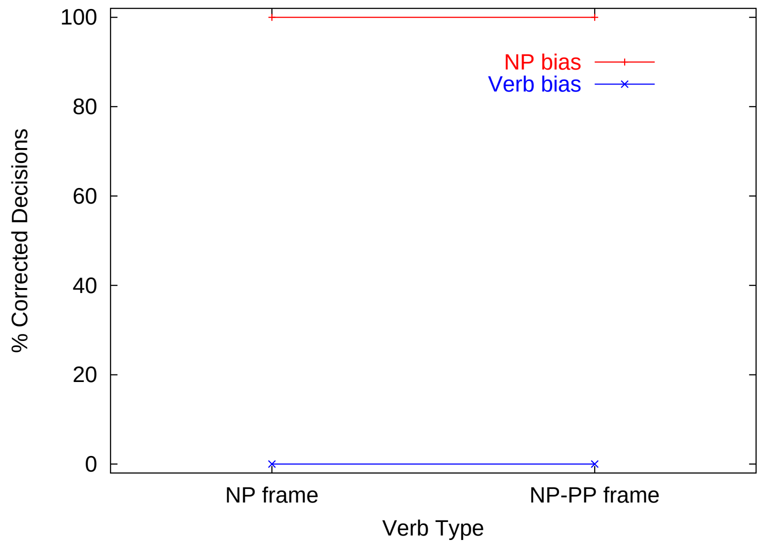
Complete Model – Results

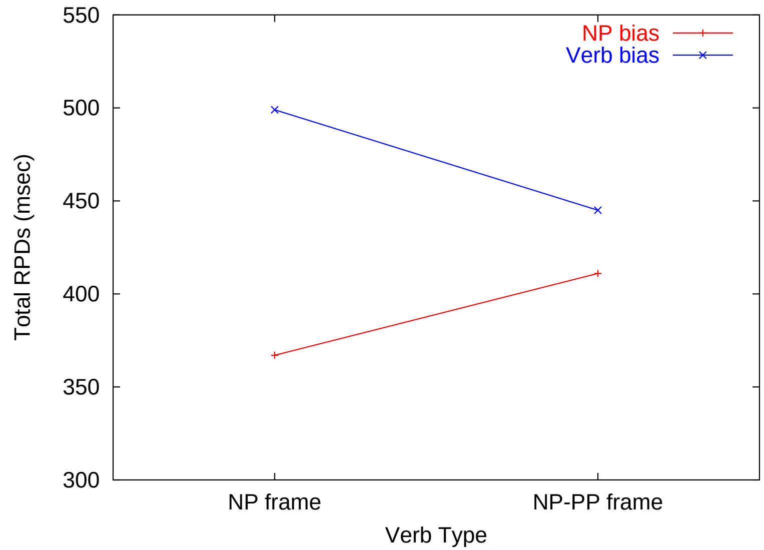




Verb second: Predictions of the model (left) compared to the Konieczny et al. (1997) data (right)

Complete Model – Results





Verb final: Predictions of the model (left) compared to the Konieczny et al. (1997) data (right)

Complete Model – Discussion

- Verb second sentences

  Replicated results for English: Verb subcategorisation influences attachment

- Verb final sentences

  Consistently wrong predictions made by the parser; this is caused by “wrong” attachment preference in the NEGRA corpus

  However, in principle, this instance of head-last processing could be covered by the model

MSc – Conclusion

- Partly successful in practice
- Even for a head-final phenomenon, successful model could be built this way in principle
- Small-scale model (tailoured to the phenomenon)

PhD – Overview

- General Idea
- Envisaged Architecture
- Sparse Data Handling
- Open Questions

General Idea

- Build a more general, larger scale model
- Model semantic processing by incrementally assigning roles to constituents returned by a parser
- Example:

  Iris annoyed the pensioner with the rock music.

  Should *with the rock music* get an instrument role from *annoyed* or should it be considered as modifying *the pensioner*?

Architecture

- Syntactic module: Parser
- Semantic module:
  - Uses syntactic hints to restrict set of possible thematic roles
  - Finds optimal role assignment for the current set of constituents
- At each step, the best parse is the one with the highest syntactic probability and the most likely role set

What goes into the role assignment model?

Consider previous example: Should *with the rock music* get a role from *annoyed*?

- Basic Model:
  - Probability of seeing the $instr$ role with the verb frame (*Semantic Subcategorisation*)
  - Probability of seeing *rock music* in the $instr$ role given the verb frame (approximates *Selectional Preferences*)

Refinements

- Probability of seeing the $instr$ role given the frame and already assigned roles (e.g. $patient$)
- Probability of seeing *rock music* in the $instr$ role given the frame and
    - *the pensioner* in the $patient$ role
    - *Iris* in the $agent$ role

Dealing with Sparse Data

- Available corpora (FrameNet/PropBank) are relatively small given the estimates I need
- If counts are unavailable, back off to simpler model
- Use noun classes instead of lemmas
  - WordNet classes
  - Clustering (e.g. Soft Clustering)

Open Questions

- How should the modules interact to be plausible?
- Which features/probabilities are plausible?
- How does evaluation work?

  E.g. Model studies, model reading times in reading time corpus