Broad Coverage Statistical Parsing with Minimalist Grammars

Reconnecting Theoretical Linguistics with Performative Computational Linguistics

---

First MIT Workshop on Minimalist Parsing
October 10-11th 2015

John Torr, Mark Steedman, Shay Cohen, Edward Stabler

A Challenge to the Minimalist Community…

THE CHALLENGE

We challenge someone to produce, by May of 2008, a working P&P parser that can be trained in a supervised fashion on a standard treebank, such as the Penn Treebank, and perform in a range comparable to state-of-the-art statistical parsers.

First, the system must use P&P in a non-trivial way. So for example, using a standard machine learning algorithm to extract a statistical parser like those in existence, supplemented by a transducer that maps Penn Tree Bank structures into P&P annotations would not satisfy the challenge. For a system to qualify, it would have to be the case that the P&P component is an integral part of the learning mechanism. Removing the P&P component must seriously degrade performance.

Second, the particular choice of parameters and their possible settings, must conform to some recognized version of a P&P theory proposed in the literature. We recognize that it may be necessary to augment the

Sproat and Lappin 2005

Background

"Minimalist Grammars…have not often been put to use in probabilistic settings…[our] improved parametrization opens up new possibilities for probabilistically-based empirical evaluation of MGs as a cognitive hypothesis about the discrete primitives of natural language grammars, and for the use of MGs in applied natural language processing."

Hunter and Dyer 2013

"Minimalist Grammar derivations can be coded by the sequence of lexical entries as it appears along the fringe of any well-formed derivation tree…this…unique readability property…might be combinable with other knowledge in a natural language understanding system [or] transfer-based machine translation system."

Stabler and Hale 2005

The role of statistics in performance parsing

- This project takes a non-emergentist view of statistical grammar induction – we still need symbols!
- Incorporating statistics simply allows the system to resolve ambiguity probabilistically. E.g:

I cut the [butter with a knife]	Very unlikely
I [[cut the butter] with a knife]	Very likely

- "For performance models, there has never been any question about the role of statistical data, also discussed in the earliest work".

  – Chomsky (2013 - taken from Reddit)

Motivation

- Construct a theory of both competence and performance.
- Building a model which scales and can handle real performance data may lead to new insights into the theory (cf Ed Stabler's talk yesterday).
- Probabilistically-based empirical evaluation of MGs (Hunter and Dyer 2013).
- Statistical parsing community can more directly exploit the rich theoretical and descriptive GB/Minimalist research.
- Potential benefits for difficult NLP tasks such as machine translation and natural language understanding.
- Principles may reduce the amount of supervision required for training (distantly supervised or even approach fully unsupervised – Johnson 2013).
- Prove Sproat and Lappin wrong!

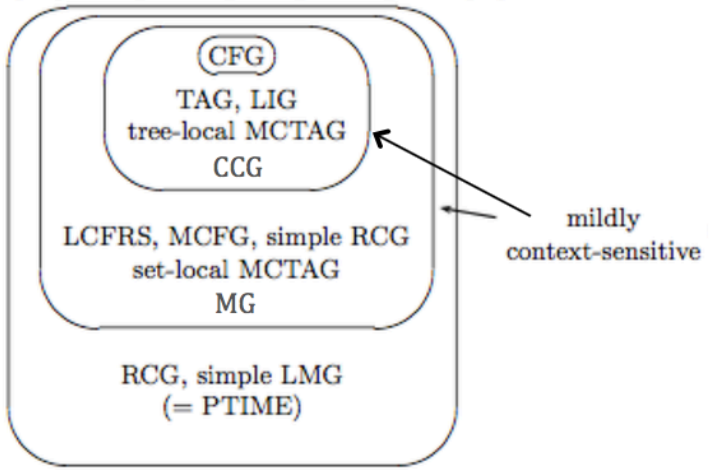
The goal

- **Create a broad coverage statistical Minimalist parser trained on a treebank. The parser should:**
    1. **Make non-trivial use of some version of Minimalism proposed in the literature (1993 – 2015).**
    2. **Handle movement directly.**
    3. **Perform efficient parsing in polynomial time.**
    4. **Perform comparably with state-of-the-art statistical parsers on recovery of local dependency relations.**
    5. **Assign expressive structures to the more interesting linguistic constructions.**
    6. **Perform competitively against other expressive parsers (CCG, TAG, HPSG) on specific tasks such as recovery of wh-antecedent-trace relations.**

The framework

Minimalist Grammars (Stabler 1997 and subsequent extensions)

- Formal, (relatively) unified, computational framework.
- Efficient parameter estimation (e.g. Hunter and Dyer 2013; Johnson 2013).
- Constrained and highly succinct formalism - mildly context sensitive (LCFRS).
- Polynomial time.



A language hierarchy of grammar formalisms (Adapted from Kallmeyer 2010:39)

Data

- All statistical models must be trained on data.
- For supervised parsing this means a corpus of parsed sentences – a treebank.
- Penn Treebank: a corpus of 49,208 sentences (952,965 words) from the Wallstreet Journal.
- Roughly adheres to the Extended Standard Theory.
- Was converted semi-automatically into CCGbank (Hockenmaier and Steedman 2007)
- Can we do the same for Minimalism?

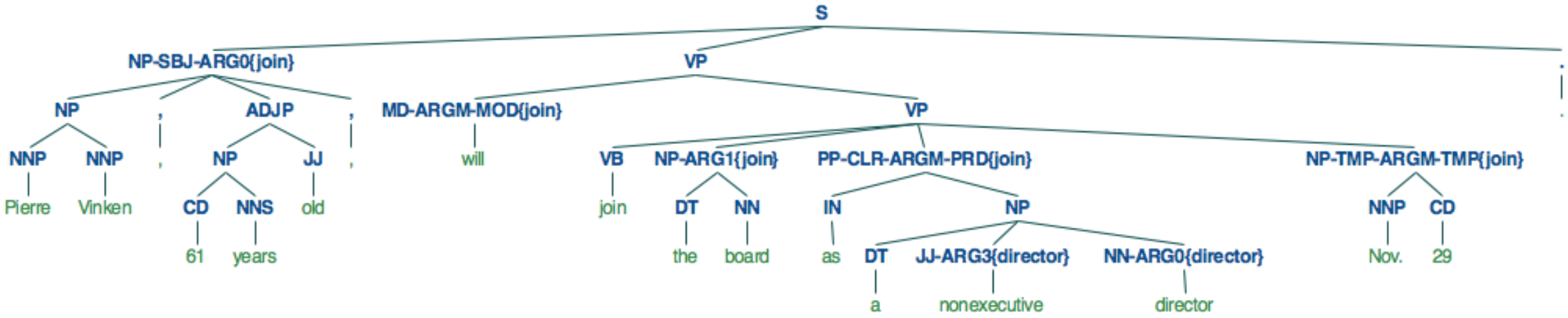
MGbank

The task:

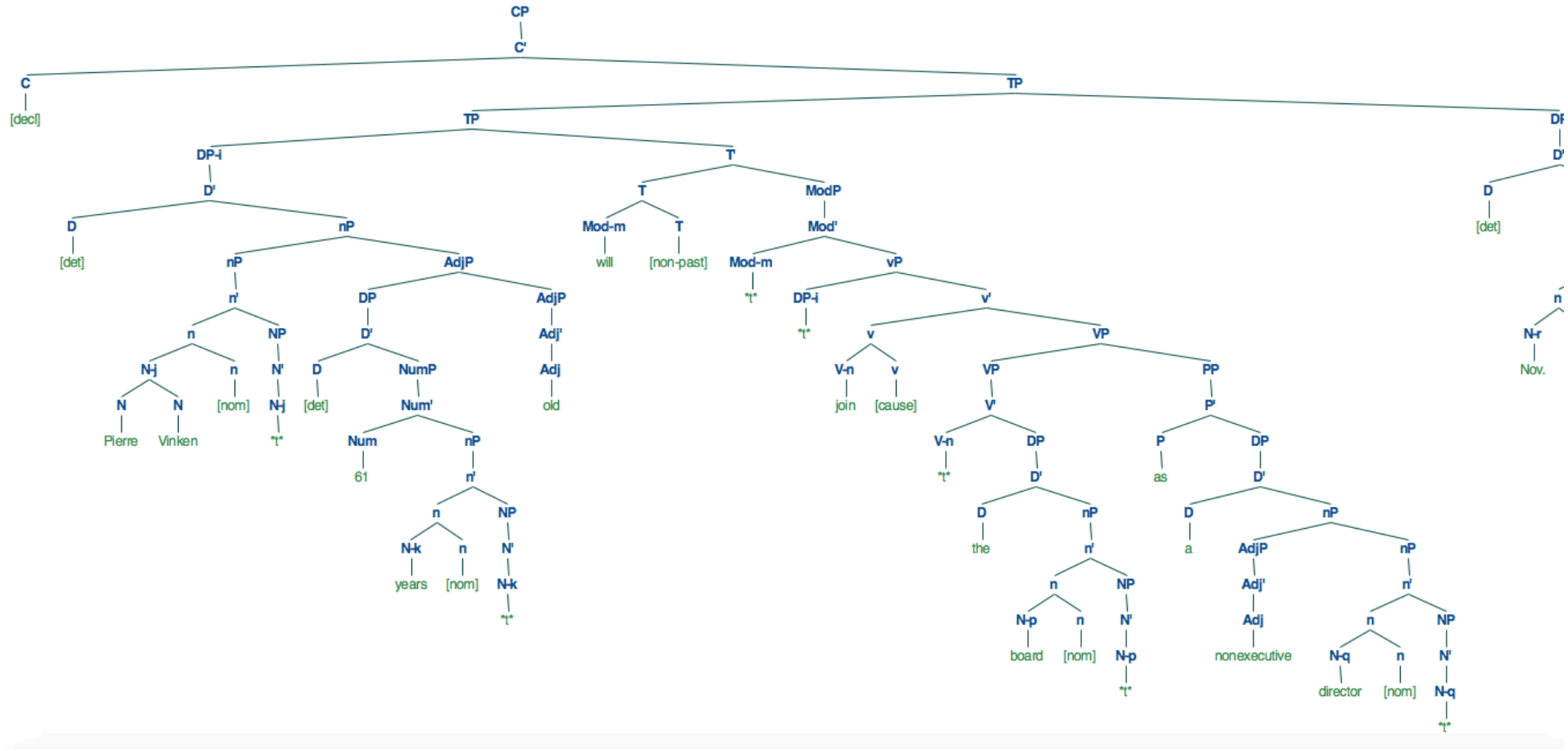
- **Construct an algorithm to (semi-) automatically convert PTB trees into MG trees.**

- **This is a hard problem! EST trees and Minimalist Trees differ considerably. For instance, sentence 1 of the PTB looks like:**

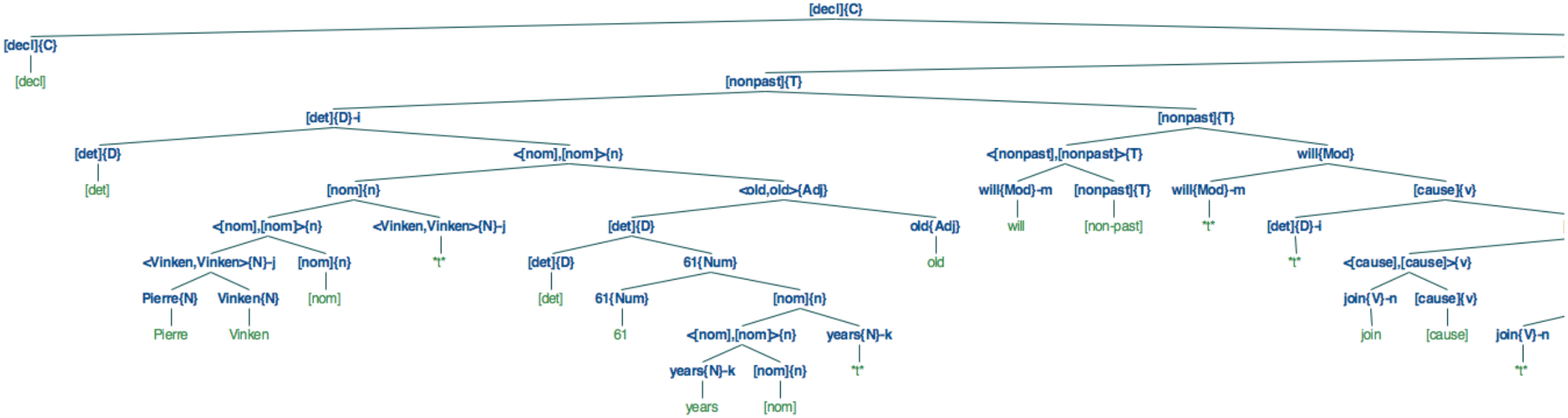
MGbank: PTB tree



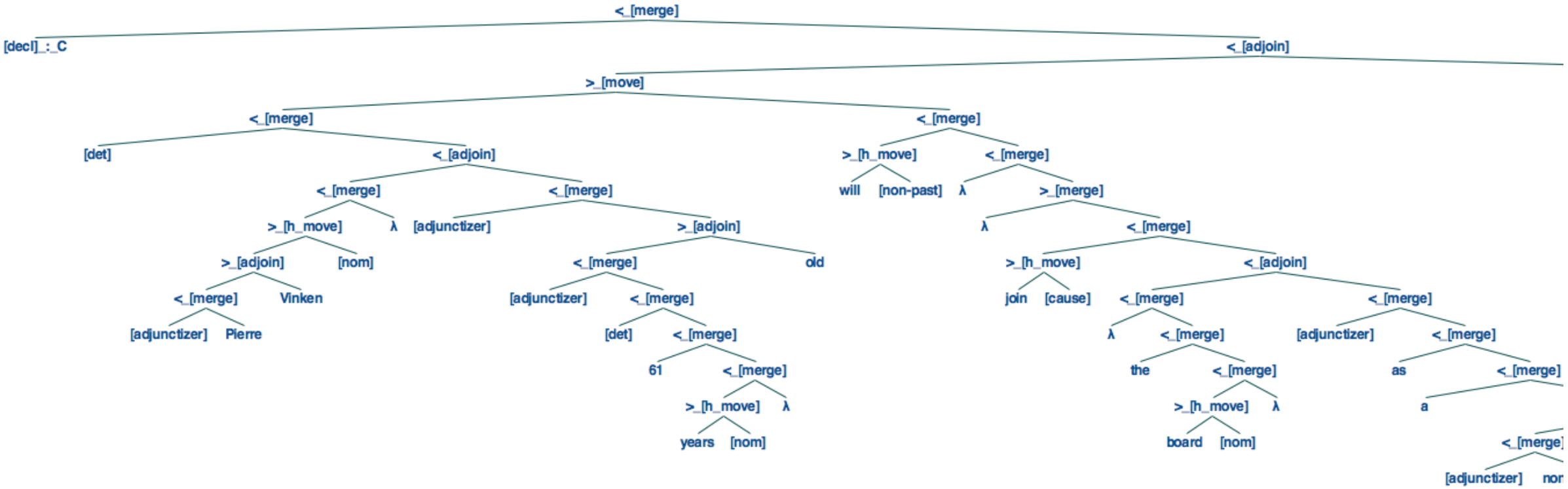
MGbank: Xbar Theory tree



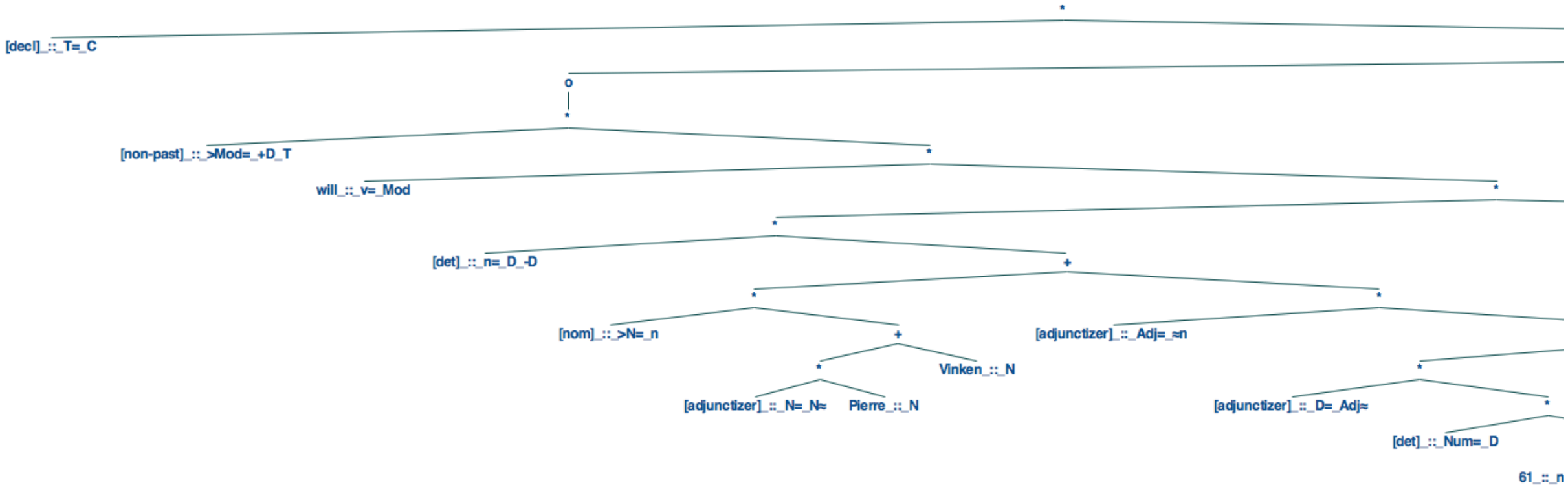
MGbank: Traditional BPS (Chomsky 1995) tree



MGbank: MG derived tree



MGbank: MG derivation tree



Difficulties

- More articulated clausal/nominal structure.
- Binary branching.
- Xbar/bare-phrase structure.
- Different government relations (functional heads).
- Additional traces (e.g. VPISH, head movement, additional adjunct traces).
- PTB does not distinguish e.g. raising and ECM from subject and object control respectively.
- Lots of rare (linguistically interesting) constructions in the tail (Zipf's Law).
- Noun Phrases very flat in the standard Penn Treebank.
- Constituents appear at different levels of structure in PTB vs MGbank.
- Complement and adjunct distinction not well defined in PTB.
- Minimalism is a research program, not a unified theory.

Additional Corpus Resources used for MGbank

Propbank (Palmer et al. 2005; University of Colorado): Added propositional and semantic role labelling information for verbs to the PTB. Useful for VPISH, UTAH, control/raising, and complement/adjunct.

Nombank (Meyers et al. 2007; New York University): Similar to Propbank, but for nominals.

Adding Noun Phrase Structure to PTB (Vadas and Curran 2007): Largely eliminates the flat nominal structure in the PTB.

The conversion algorithm

- Does not operate on the Penn Trees directly.
- Instead, as the Penn tree is traversed, certain nodes trigger the generation of an empty Minimalist *extended projection* (Grimshaw 2000).
- This empty structure is then 'filled in' with heads, and arguments and adjuncts, the latter themselves extended projections.
- Coordination is handled slightly differently.
- Post-processing involves extraposition movements for e.g. heavy NP-shift.
- Finally, the tree is compressed, removing unfilled projections.

After 6 months…

- **Well over 90% of the Penn Treebank structure has been converted into MG structure.**
- **52% of trees have all words present and correctly ordered, 48% have on average 1 error but are still useful for statistical modelling.**
- **Sounds quite good, BUT! Now into the tail..**

Conclusion

- MGbank is coming along nicely but still a lot to do!
- Will almost certainly require some hand annotation at some point.
- Volunteers will be very welcome! (john.torr@cantab.net)
- Thank you MIT!