A Psycholinguistically Motivated Version of TAG

Vera Demberg and Frank Keller

School of Informatics
The University of Edinburgh

TAG+ 9 Workshop, June 7th 2008

Introduction

Motivation

- To build a computational model of human sentence processing which is **psycholinguistically plausible**
- Model of **comprehension** (as opposed to production)
- **Assumptions:** incrementality, connectedness, prediction

Goals

- Better model of human sentence processing
- Incremental processing beneficial for models of speech recognition, speech-to-speech translation

The Sentence Processing Model

The Model

- Consists of a **parsing process** and a **linking theory**
- Key requirements that the model should fulfill:
  - **Incrementality:** Word-by-word processing, eager integration of each word into a syntactic structure
  - **Connectedness:** Most strict version of incrementality – all words in a sentence are connected into a single structure
  - **Prediction:** Humans predict upcoming structure and lexemes before encountering them
- The parser should therefore also implement these properties

Empirical Evidence for Incrementality and Prediction

Coordination processing: **structural binding** in c-command relation [Sturt & Lombardo 2005]

Experimental Findings: Incrementality & Connectedness

1. The pilot embarassed John and put himself in an awkward situation.
2. The pilot embarassed Mary and put herself in an awkward situation.
3. The pilot embarassed John and put him in an awkward situation.
4. The pilot embarassed Mary and put her in an awkward situation.

Gender default mismatch difficulty occurred at first pass reading on pronoun in condition (2) where herself is c-commanded by “pilot” but not in condition (4).

Empirical Evidence for Incrementality and Prediction

Visual world experiment: **anticipatory eye-movements** show that people predict subsequent input [Kamide et al. 2003]

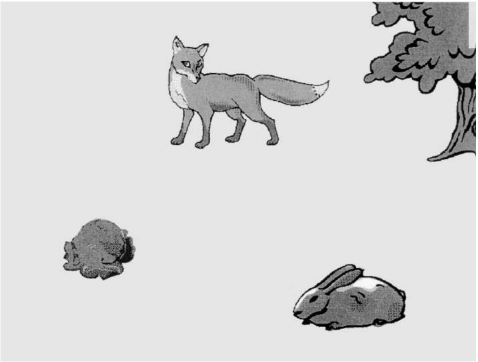
Experimental Findings: Incrementality and Prediction

"Der Hase frisst gleich den Kohl."

The Hare-nom will eat soon the cabbage-acc.

"Den Hasen frisst gleich der Fuchs."

The Hare-acc will eat soon the fox-nom.



Empirical Evidence for Incrementality and Prediction

Either...or processing:
processing facilitation through prediction [Staub & Clifton 2002]

Experimental Findings: Prediction

- either... or
  presence of “either” leads to shorter fixation times on “or” and the second conjunct
- general treatment of two-part constructions [Cristea & Webber, 1997]
- syntactic parallelism
  Second conjunct processed faster if internal structure identical to first conjunct [Frazier et al., 2000]

Choice of Grammar Formalism

- Chose **LTAG** as a basis
- **Extended domain of locality** gives powerful tool for implementing prediction (so phrase-structure grammar not suitable)
- **Incrementality and Connectedness** easier to realize than with e.g. CCG
  - no incremental derivation for object relative clauses in CCG
  - incrementality in coordination problematic in CCG

Overview

Other versions of Incremental TAG / Incremental Parsers

Previous work on incremental TAG parsing:

- **Spinal LTAG** [Shen and Joshi, 2005]
  - not connected
  - no subcategorisation information
- **Incremental TAG** [Kato et al., 2004]
  - trees are leftmost expanded
  - no proper modifier / argument distinction
- **Dynamic Version of TAG** [Mazzei et al., 2007]
  - most similar, but different grain sizes for prediction

Previous work on incremental parsing:

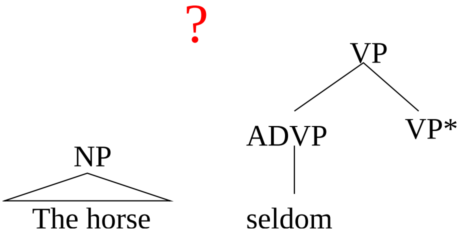
- PCFG [Roark, 2001]
- Dependency Parser [Nivre, 2004]

The Interplay of Connectivity and Prediction

- Connectivity and Prediction interact closely
- We need a fully connected structure in order to determine what's predicted (i.e. what is expected in order to build a grammatical sentence?)
- We need prediction in order to achieve connectivity
- Example of how parsing process works in PLTAG for sentence “The horse seldom won a prize”.

The Interplay of Connectivity and Prediction (2)

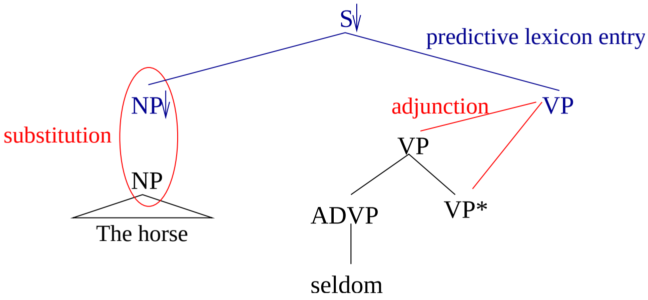
Normal LTAG does not allow for **connectedness**



- Elementary trees cannot always be connected directly

The Interplay of Connectivity and Prediction (3)

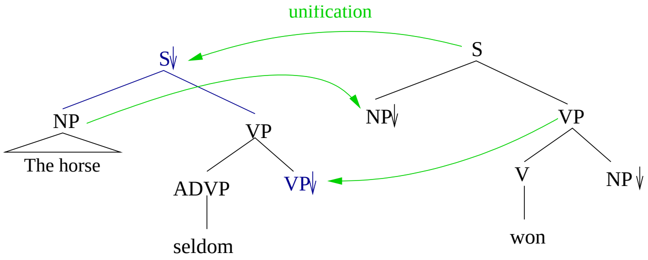
Extend lexicon to include **predictive entries**



- Insert connecting structure
- Predicted nodes marked by "↓"
- Structure is non-lexicalised "predictive lexicon entry"

The Interplay of Connectivity and Prediction (4)

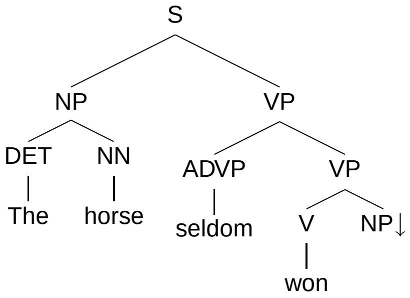
Introduce a **Verification** step



- Observe **accessibility** constraints and **dominance** relations
- Valid sentence analyses must not contain any open predicted nodes which have not been verified

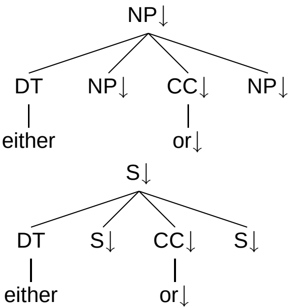
Prediction from the Lexicon

Substitution nodes that are to the right of a lexical anchor of an elementary tree typically generate predictions.

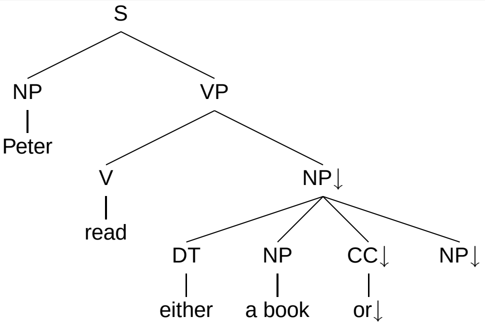


- Exploit **extended domain of locality** to design lexicon entries in order to model psycholinguistic findings such as prediction in "either...or" constructions.

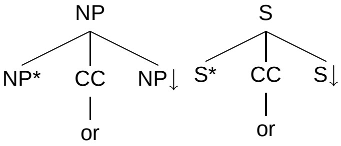
Extended domain of locality



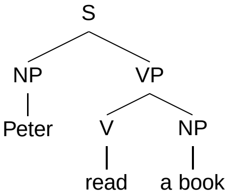
(a) lexicon entry for *either*



(b) with *either*



(c) lexicon entries for *or*



(d) no *either*

Determining Elementary Trees

- **Converted Penn Treebank** into TAG format [Xia et al. 2000]
- **Head percolation table** for determining how to cut up the tree into elementary trees [Magerman 1994]
- **PropBank** [Palmer et al. 2003] and **NomBank** [Vadas & Curran 2007] for discriminating arguments and modifiers
- Determine domain of locality (e.g. either... or, pick... up)

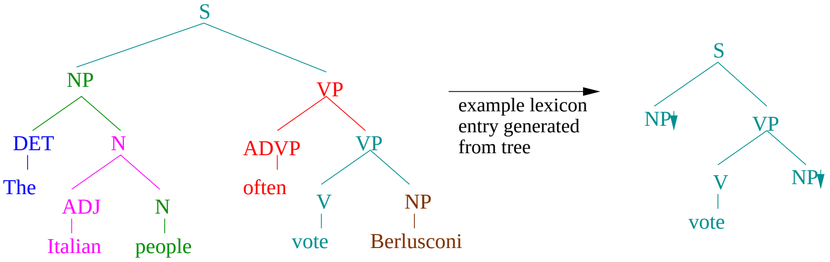
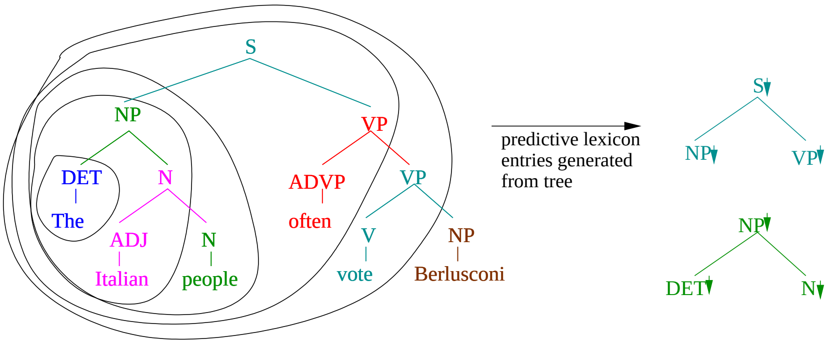


Figure: Generating lexicon entries from the Penn Treebank for an example sentence

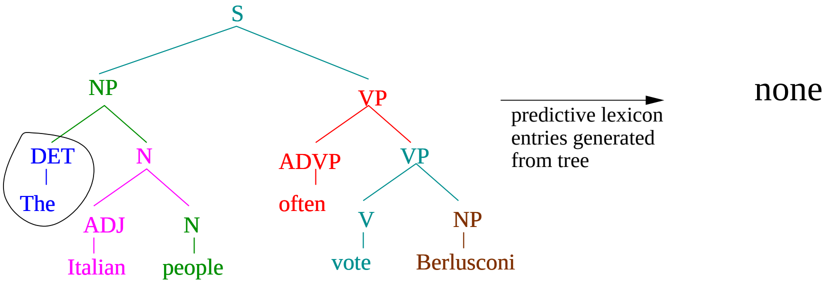
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



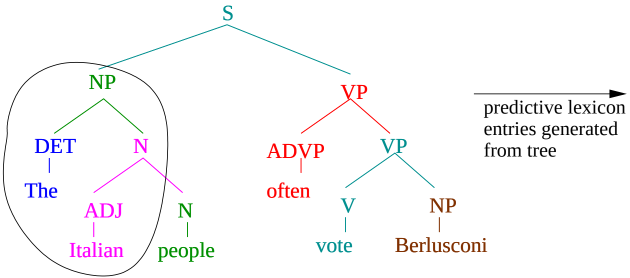
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



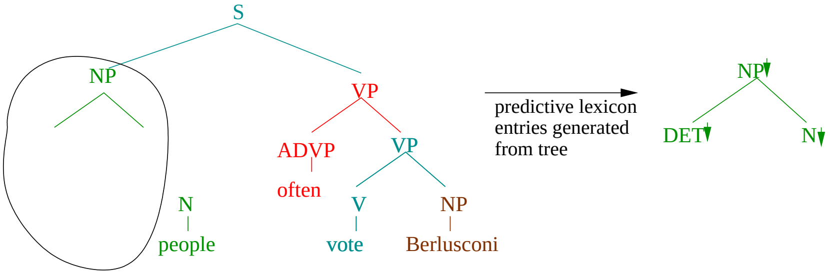
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



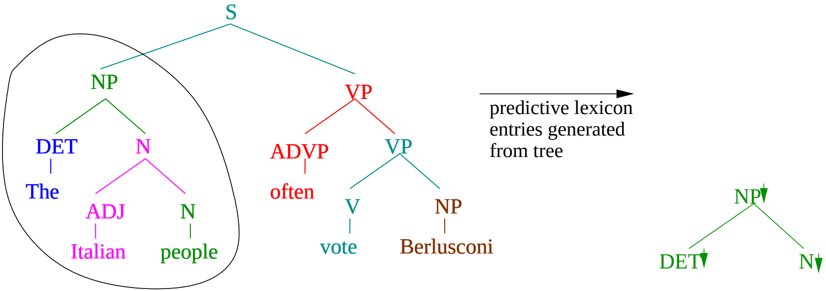
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



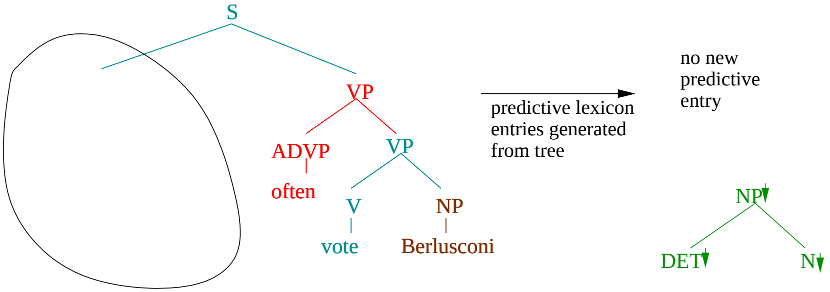
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



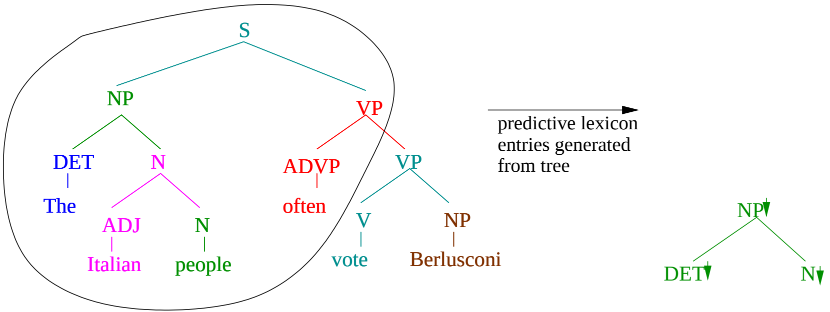
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



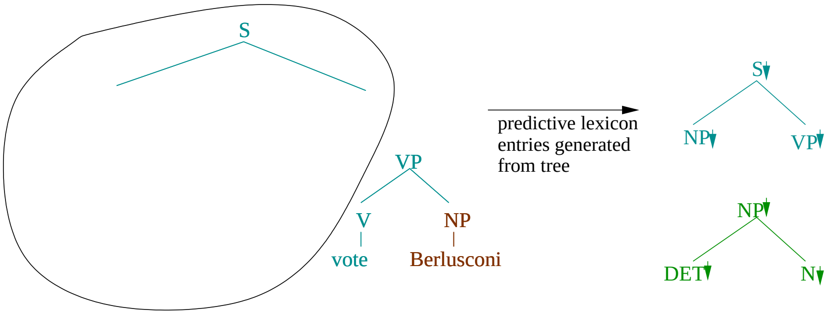
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



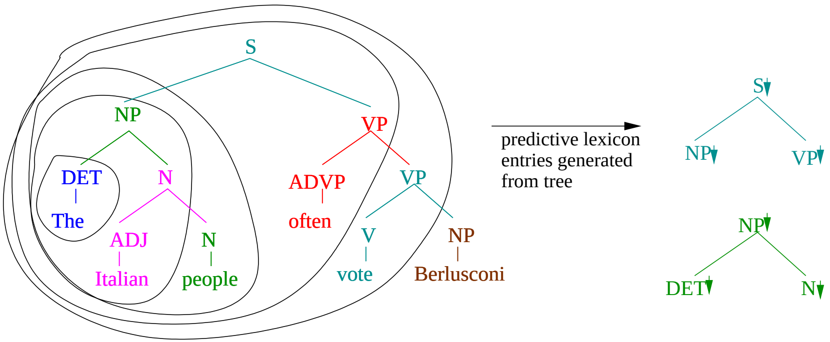
Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



Connection Paths

- Connection path: minimal amount of structure that is needed to connect words $w_1 \cdots w_n$ under one node [Lombardo & Sturt, 2002]
- Determine predicted structures necessary by connectivity



The Linking Theory

- Given an incremental TAG parser, we can analyse sentence, and know where e.g. a lot of prediction is required, where a lot of nodes need to be matched up etc.
- To predict processing difficulty, need to correlate the processes of the parser to some measure of difficulty

Ingredients of Linking Theory

Basic effects that we want to capture, and which have previously been shown to be significant predictors of reading time [e.g. Gibson, 1998; Hale, 2001; Lewis & Vasishth, 2005; Demberg & Keller, 2007]

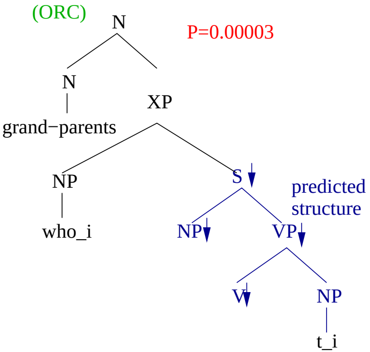
- **Surprisal effects**
  unexpected material is more difficult to process than predicted material
- **Locality effects**
  long distance relationships as in center embedding are more difficult to process
- **Activation and Memory effects**
  syntactic rules and lexical items are easier to process when they have a high activation level

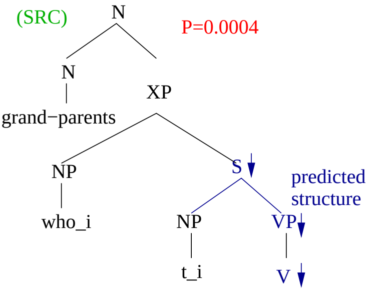
Linking the Parsing Process to Sentence Processing Difficulty (work in progress)

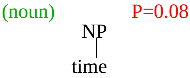
- At each stage, we have a set of **expectations** E of what is needed to build a grammatical sentence (incrementality with full connectivity)
- **Ranked parallelism**: the expectations are weighted according to the **Probability** $P(e)$ of the analysis that generated them
- Expectations have a timestamp t for when they were generated or last activated, and are held in **memory**
- **Decay** f: the longer ago an expectation was generated, the more difficult it is to retrieve it; reflects the activation level
- **Verification** causes processing difficulty when
    1. Expectation satisfied (noun expected – noun found) E_i
    2. Analysis that generated expectation incompatible with new input E_d

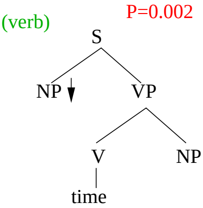
$$\text{processing difficulty } D_w \propto \sum_{e \in E_i} f\left(\frac{1}{P(e)}\right) + \sum_{e \in E_d} f(P(e)) \qquad (1)$$

An example

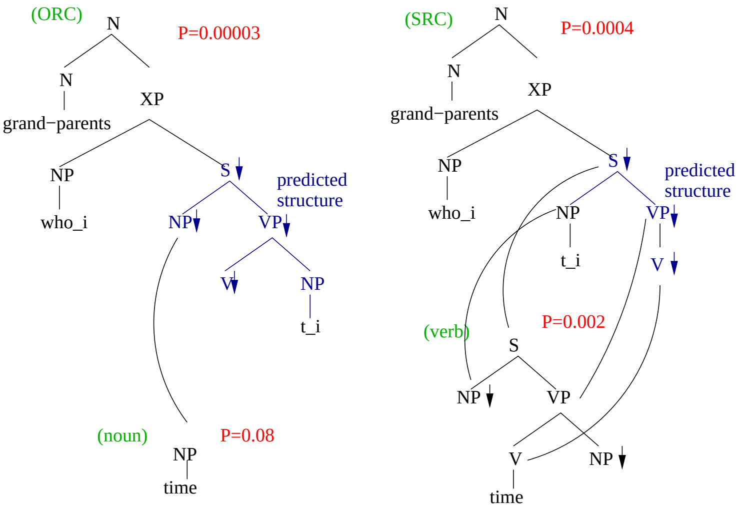

(ORC)
P=0.00003
N
N
XP
grand−parents
NP
S
who_i
NP
VP
predicted
structure
V
NP
t_i


(SRC)
P=0.0004
N
N
XP
grand−parents
NP
S
who_i
NP
VP
predicted
structure
t_i
V


(noun)
P=0.08
NP
time


(verb)
P=0.002
S
NP
VP
V
NP
time

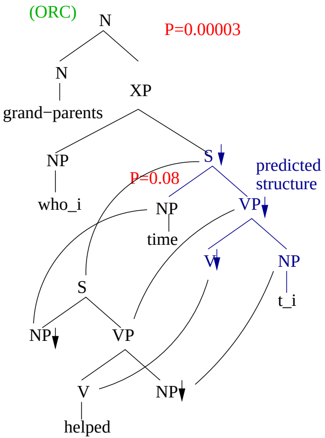
An example



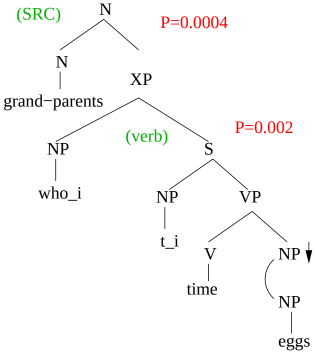
1 integration
timestamp: 1 time unit

4 integrations
timestamp: 1 time unit

An example



5 integrations:
4 @ timestamp 2
1 @ timestamp 1



1 integration
timestamp 2

compare: integration at verb in SRC was 4 @ timestamp 1

integration at ORC noun was 1 @ timestamp 1

Conclusions

Summary

- Motivated why it is interesting to model incrementality with full connectivity, and prediction based on recent findings in psycholinguistics
- Suggested an incremental TAG version that implements these requirements
- Proposed a linking theory that relates the parsing process to a theory of human sentence processing

Future Work

- Define a **probability model** for PLTAG
- Implement an **incremental parser**
- **Evaluation** of processing difficulty predictions on eye-tracking corpus and comparison to other theories of sentence processing

Thanks for your attention!

QUESTIONS?

TAG vs. CCG – Incrementality

CCG is less easily incrementalizable than TAG.

- Coordination [Lombardo & Sturt, 2005]

Example

the pilot	*embarrassed Mary*	*and*	*put herself*	*in an awkward situation*
$S/(S\backslash NP)$	$S\backslash NP$	$CONJ$	$(S\backslash NP)/PP$	PP

$(S\backslash NP)/PP \;\; PP \Rightarrow S\backslash NP \quad (>)$

$S\backslash NP \;\; CONJ \;\; S\backslash NP \Rightarrow S\backslash NP \quad (<\Phi>)$

$S/(S\backslash NP) \;\; S\backslash NP \Rightarrow S \quad (>\mathbf{B})$

Figure: Binding would only occur after full processing of second conjunct according to CCG derivation. However, the empirical finding is that humans experience difficulty of gender mismatch as soon as they hit the reflexive pronoun.

TAG vs. CCG – Incrementality

CCG is less easily incrementalizable than TAG.

- Object Relative Clauses

Example

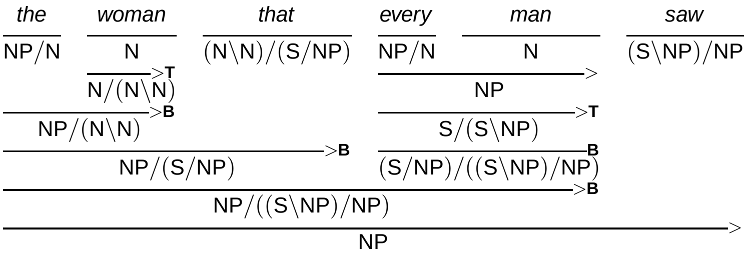


Figure: Example of incrementalized derivation for object relative clause in CCG. It is not possible to make a fully incremental version inside the ORC NP “every man”.

TAG vs. CCG – Symmetry in Modification

Pre- and post-modification in CCG

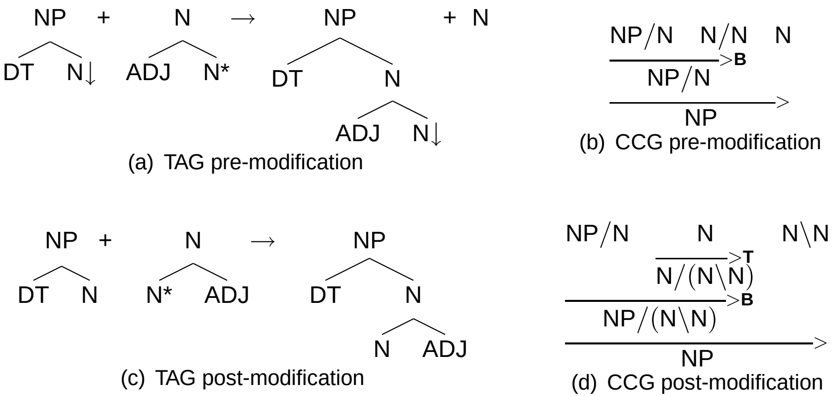


Figure: Comparision of pre- and post-modification in TAG and CCG

TAG vs. PCFG

Phrase structure Grammar

- Standard arc-eager parsing completes all rules when the last instance of the innermost rule is found
  VP $\longrightarrow$ V • PP
  PP $\longrightarrow$ P • NP
  NP $\longrightarrow$ DT • N
- Composition in parsing process would be necessary
  VP $\longrightarrow$ V [P [DT • N]] [Thompson et al. 1991]
- Same amount of prediction would be needed
- No extended domain of locality