Incremental, Predictive Parsing with Psycholinguistically Motivated Tree-Adjoining Grammar

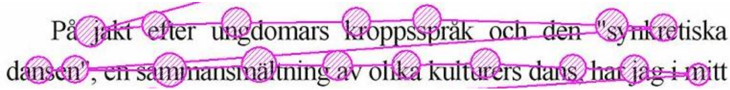
Frank Keller

Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh
http://homepages.inf.ed.ac.uk/keller/

Joint work with Vera Demberg and Alexander Koller

Incrementality

Text and speech are perceived *serially.*



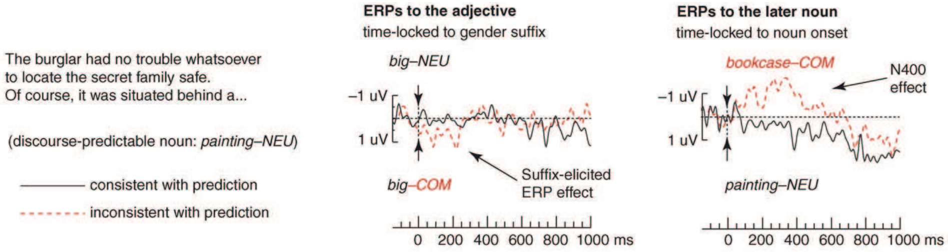
Human language processing is adapted to this: sentence comprehension proceeds *incrementally:*

- the interpretation of a sentence is built word by word;
- each new word is integrated as fully as possible into a representation of the sentence thus far;
- processing effort depends on the properties of the word and its relationship to the preceding context.

Discourse Prediction

Not only is processing word-by-word, it is also *predictive:* comprehenders anticipate upcoming linguistic material.

van Berkum et al. (2005) show that contextual information is used to predict specific lexical items; processing difficulty ensues if input is incompatible with the prediction (ERP study).



Structural Prediction

Staub & Clifton (2006) show that the sentence processor can also make *structural predictions:*

(1) Peter read either a book or an essay in the school magazine.

(2) Peter read a book or an essay in the school magazine.

The presence of *either* leads to shorter reading times on *or* and on the NP that follows it (eye-tracking study).

The word *either* makes it possible to anticipate an upcoming NP conjunction (rather than VP conjunction).

Semantic Prediction

Visual world paradigm:

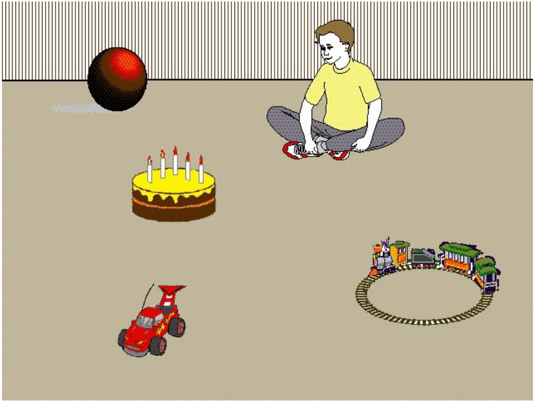
- image and speech presented synchronously;
- eye-movements reflect listeners' interpretation of input;
- they can also indicate predictions about upcoming input.

Altmann & Kamide (1999) use this paradigm to provided evidence for *semantic prediction*. They presented sentences such as:

(3) a. The boy will eat . . .
b. The boy will move . . .

together with a scene that contained one edible but several movable objects.

Semantic Prediction



When participants heard *eat*, they looked more at the cake.

Evidence for prediction induced by semantic restrictions of the verb.

Granularity of Prediction

What is the *granularity of prediction?* We saw predictions can be triggered by:

- discourse context;
- specific collocations (*either ... or*);
- semantic restrictions of a lexical item.

But can we get predictions from *lexically specific syntactic* information?

We will look at an experiment in detail that shows prediction based on *verb subcategorization* (joint work with Manabu Arai).

Design

Compare obligatory transitive verbs (e.g., *offend*) and intransitive verbs (e.g., *frown*):

(4) a. All of the sudden, the inmate offended the judge.
  b. All of the sudden, the inmate frowned at the judge.
  c. All of the sudden, the inmate frowned and the judge threw the gloves.

Listeners' looks at the verb indicate which verb frame they assume.

Materials

(5) All of the sudden, the inmate *offended* the judge.

Listeners predict upcoming patient information on hearing the verb and look at the judge.

Materials

(5) All of the sudden, the inmate *frowned* at the judge.

No prediction on hearing the verb. But listeners predict upcoming patient information on hearing *at* and look at the judge.

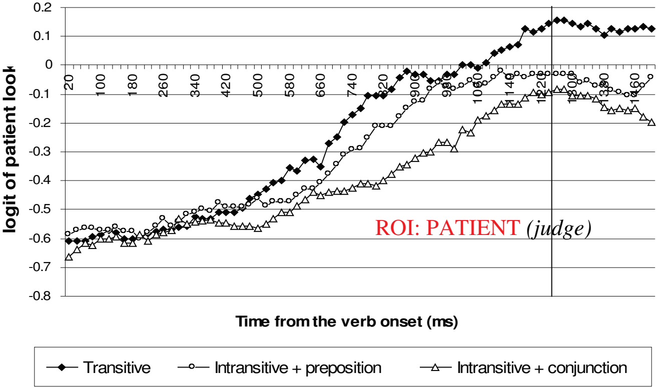
Materials

(5) All of the sudden, the inmate *frowned* and the judge threw the gloves.

No prediction on hearing the verb and *and*.

Results

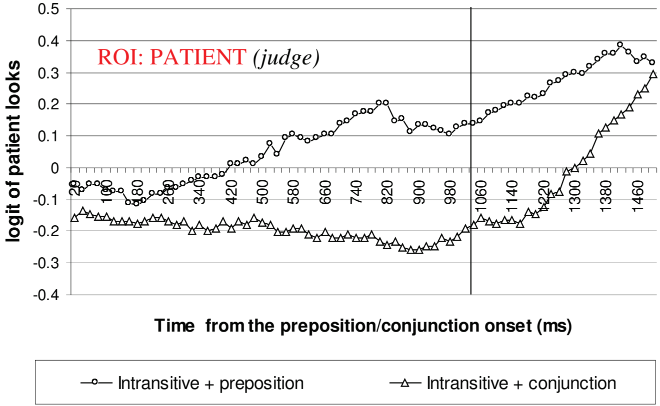
Gazes after the verb onset:



Participants looked more at a patient picture on hearing transitive verbs (*offended*) than intransitive verbs (*frowned*).

Results

Gazes after hearing *at* or *and*:



Participants predicted and looked more at a patient picture on hearing *at* than on hearing *and*.

Summary

Evidence for the use of verb-specific *subcategorization information in prediction:*

- participants predicted a direct object following a transitive verb more than following an intransitive verb;
- they made a similar predictions at the preposition following an intransitive verb;
- but not if a conjunction followed the intransitive verb.

Conceptual Issues

Challenge: develop a model of prediction in sentence processing that accounts for these experimental results. Assumptions:

- structures are built incrementally (word by word);
- partial structures do not contain unconnected nodes;
- upcoming syntactic material is predicted.

Evidence for *connectedness:* Sturt & Lombardo (2005). Existing incremental parsers don't build fully connected structures.

Our approach: devise a *grammar formalism* that supports incrementality and connectedness; prediction then follows.

Implementing Prediction

Experimental results inform our model regarding the *granularity of prediction*. The model predicts:

- lexical items when they are syntactically required (e.g., *either . . . or*, *pick . . . up*);
- syntactic structure when required by subcat frames;
- syntactic structure when required by connectedness.

Formalism

We propose Psycholinguistically Motivated TAG (PLTAG), a variant of *tree-adjoining grammar:*

- in standard TAG, the lexicon consists of initial trees and auxiliary trees (both are lexicalized);
- we add unlexicalized *predictive trees* to achieve connectivity;
- the standard TAG operations are substitution and adjunction;
- we add *verification* to verify predictive trees;
- we use TAG's extended domain of locality for lexical prediction.

PLTAG supports parsing with incremental, fully connected structures (Demberg & Keller 2008).

Formalism

Lexicon:

- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

- Substitution
- Adjunction
- Verification (PLTAG)

Formalism

Lexicon:

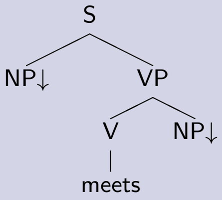
- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

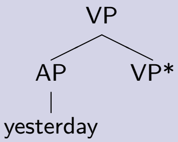
- Substitution
- Adjunction
- Verification (PLTAG)

Example

Initial Tree:



Auxiliary Tree:



Formalism

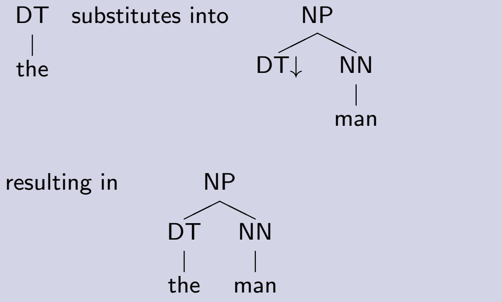
Lexicon:

- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

- Substitution
- Adjunction
- Verification (PLTAG)

Example



Formalism

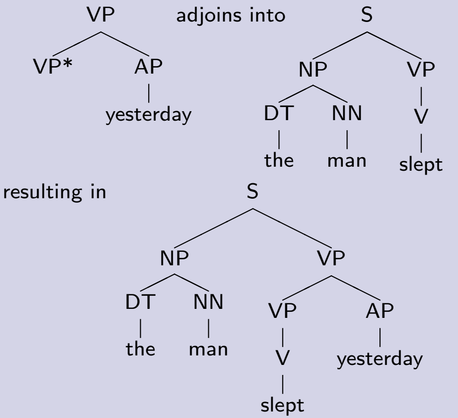
Lexicon:

- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

- Substitution
- Adjunction
- Verification (PLTAG)

Example



Formalism

Lexicon:

- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

- Substitution
- Adjunction
- Verification (PLTAG)

Example

Initial Prediction Tree:

S_s

NP^s VP^s_s

Auxiliary Prediction Tree:

VP_s

AP^s_s $VP*^s$

Index *s* marks predicted node.

Formalism

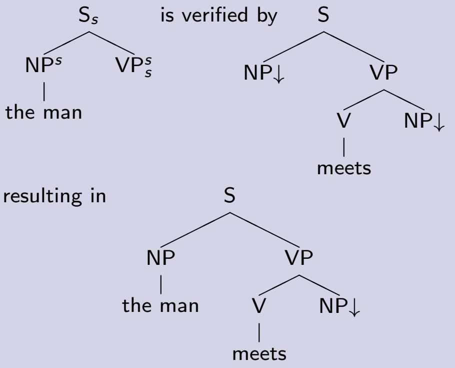
Lexicon:

- Standard TAG lexicon
- Predictive lexicon (PLTAG)

Operations:

- Substitution
- Adjunction
- Verification (PLTAG)

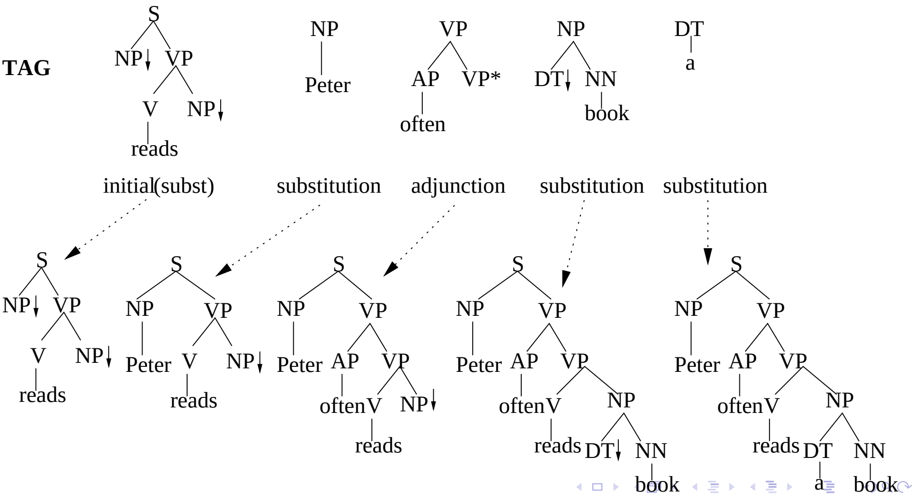
Example



All nodes indexed with s have to be verified.

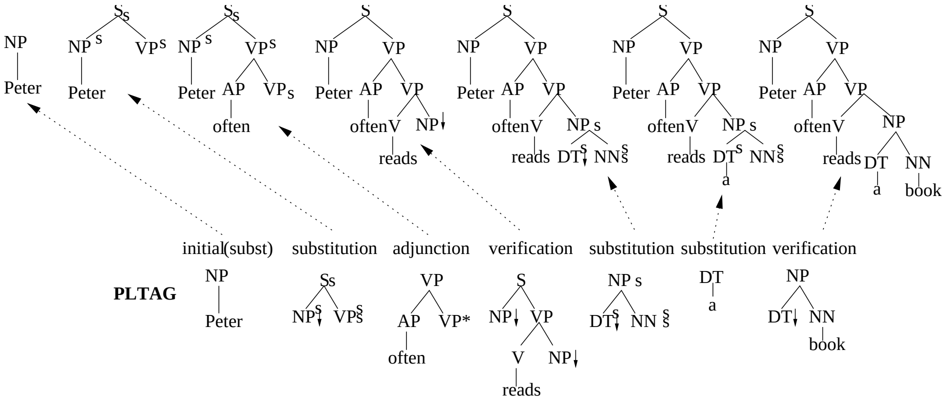
Comparison with and TAG

TAG derivations are not always incremental.



Comparison with and TAG

PLTAG derivation are always incremental and fully connected.



Modeling Prediction

PLTAG assumes three types of prediction:

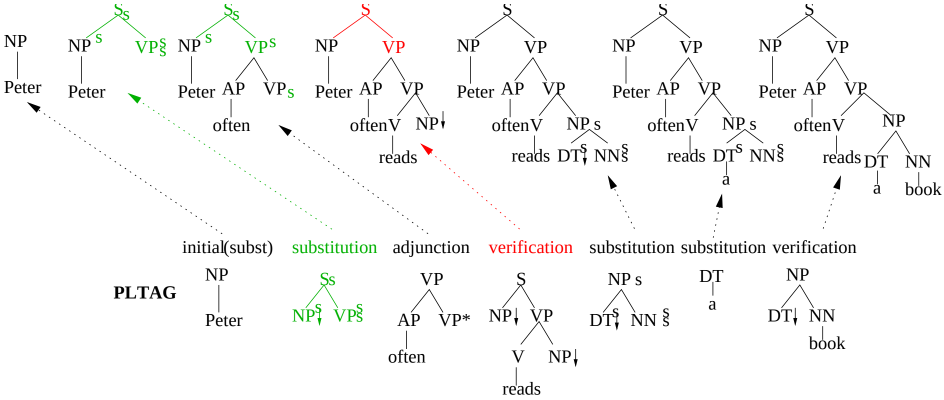
- predictive nodes (required for connectivity);
- open substitution nodes (subcategorization);
- lexical prediction (e.g., *either ... or*).

Connectedness and prediction interact closely:

- in order to achieve incrementality with full connectedness, upcoming nodes have to be predicted;
- in a fully connected structure, predictions can be read off straightforwardly (all open prediction and substitution nodes).

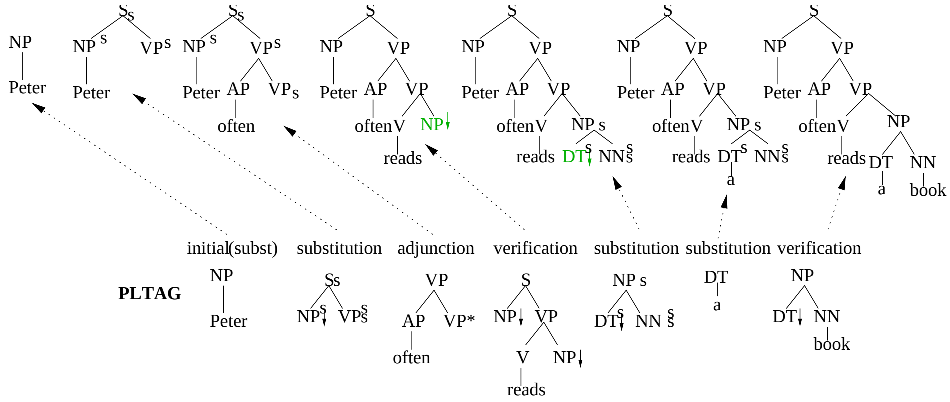
Modeling Prediction

Predictive Nodes:



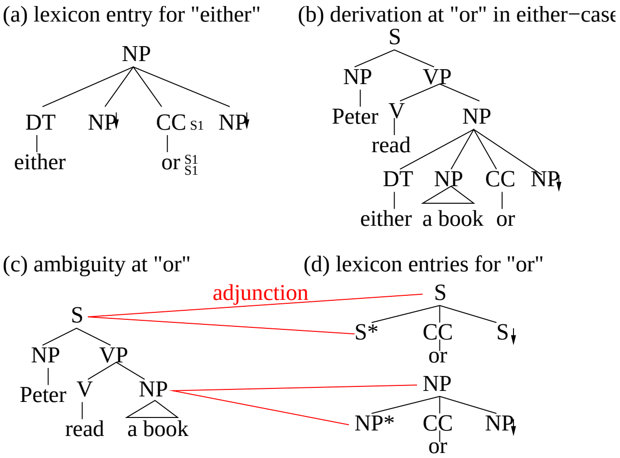
Modeling Prediction

Open Substitution Nodes:



Modeling Prediction

Lexical prediction based on TAG's extended domain of locality:



An Incremental Parser for PLTAG

In order to construct an incremental parser for PLTAG, we need to:

1. convert the Penn Treebank into PLTAG format;
2. induce a lexicon from it;
3. develop an incremental parsing algorithm;
4. devise a probability model;
5. formulate a linking theory.

Step 1: Treebank Conversion

Convert Penn Treebank into TAG format (Xia et al. 2000) using:

- *head percolation table* for determining how to cut up a tree into elementary trees (Magerman 1995);
- *Propbank* for distinguishing arguments and modifiers (Palmer et al. 2003);
- *noun phrase annotation* to derive NP-internal structure (Vadas & Curran 2007).

The resulting trees are less flat, contain head information, and argument/modifier distinction.

Step 2: Lexicon Induction

A *standard TAG lexicon* can be derived from the TAG Treebank by cutting up the trees into initial trees and adjunction trees.

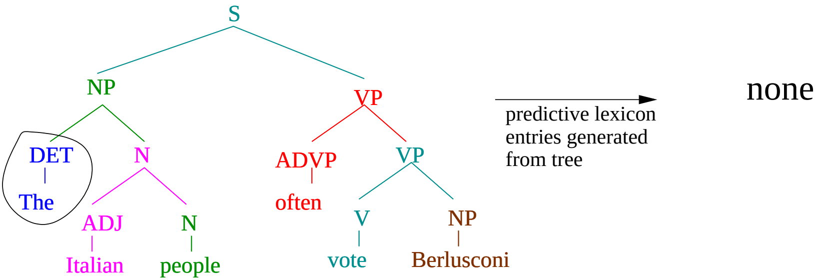
For the *predictive lexicon,* we need the notion of connection path.

Connection Path

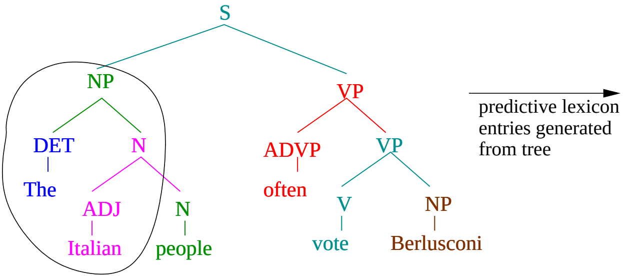
The connection path of w_1 is the minimal amount of structure needed to connect words $w_1 \dots w_i$ under one node (Sturt et al. 2003).

Essentially, we determine which parts of the tree we need to predict to achieve connectivity.

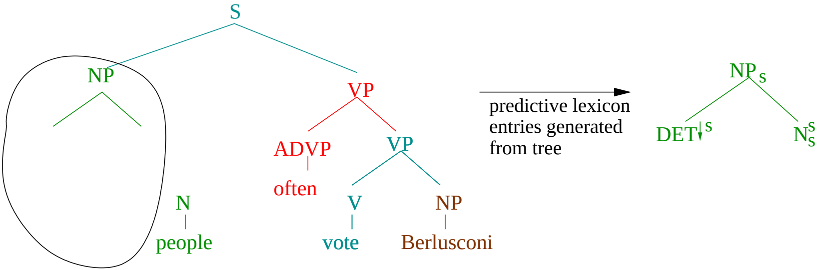
Step 2: Lexicon Induction



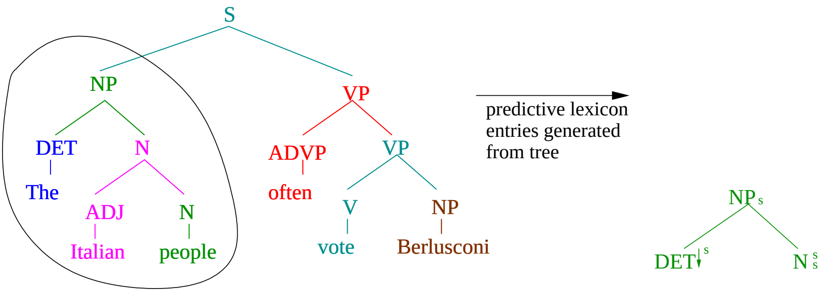
Step 2: Lexicon Induction



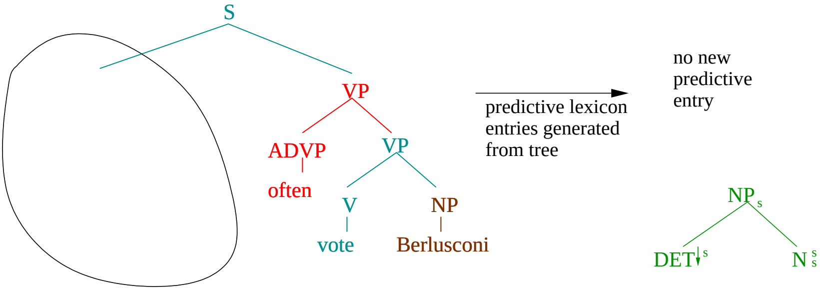
Step 2: Lexicon Induction



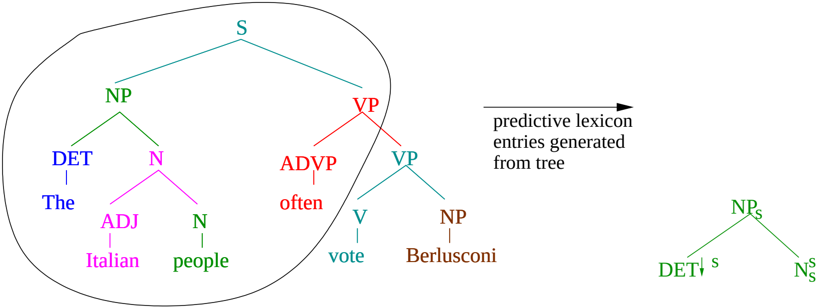
Step 2: Lexicon Induction



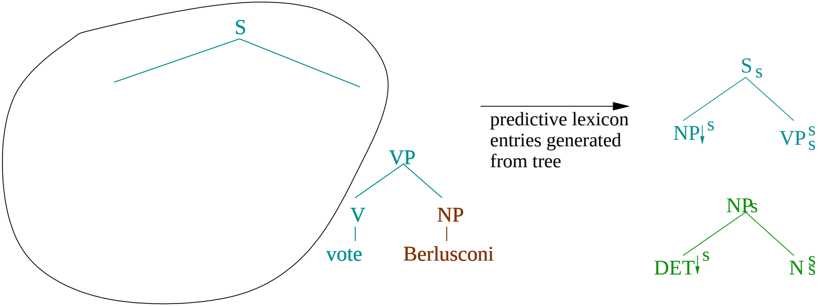
Step 2: Lexicon Induction



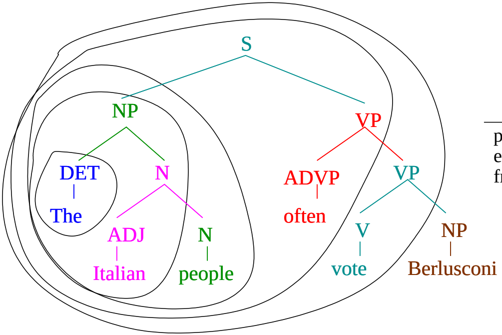
Step 2: Lexicon Induction



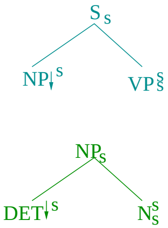
Step 2: Lexicon Induction



Step 2: Lexicon Induction



predictive lexicon entries generated from tree



Step 3: Parsing Algorithm

Properties:

- incrementally builds fully connected partial structures;
- only allows valid partial PLTAG structures;
- constructs all possible structures in parallel.

At word w_i, retrieve elementary tree ϵ for w_i and connect it to the prefix tree β for $w_1 \dots w_{i-1}$:

- parsing operations: substitution, adjunction, verification;
- dependent on status of β and ϵ: standard or predictive tree.

Step 4: Probability Model

The following properties need to hold (Chiang 2000):

Substitution: $\sum_{\epsilon} P(\epsilon|\eta_\beta) = 1$

Adjunction: $\sum_{\epsilon} P(\epsilon|\eta_\beta) + P(NONE|\eta_\beta) = 1$

Verification: $\sum_{\epsilon} P(\epsilon|\pi_\beta) = 1$

where $P(\epsilon|\eta_\beta) = P(\tau_\epsilon|\eta_\beta)P(\lambda_\epsilon|\tau_\epsilon, \lambda_\beta)$
and $P(\epsilon|\pi_\beta) = P(\tau_\epsilon|\pi_\beta)P(\lambda_\epsilon|\tau_\epsilon)$

elementary tree	ϵ	prefix tree	β	prediction tree	π
tree structure	τ	integration point node	η	a tree's head leaf	λ

Step 5: Linking Theory

The linking theory translates parser states into processing difficulty:

- elementary tree ϵ_{w_i} is integrated with prefix tree $\beta_{w_1 \ldots w_{i-1}}$;
- processing difficulty proportional to change in distribution $P(\beta)$ from w_{i-1} to w_i;
- each predicted tree π has a time-stamp t;
- at verification, decay d is calculated based on t (recently accessed structures are easier to integrate).

$$D_{w_i} = \overbrace{-\log \sum_{\beta_{w_1 \ldots w_i}} P(\beta_{w_1 \ldots w_i}) + \log \sum_{\beta_{w_1 \ldots w_{i-1}}} P(\beta_{w_1 \ldots w_{i-1}})}^{\textit{Surprisal}} \left. - \log \sum_{\pi} P(\pi)^{(1-d^{t_\pi})} \right\} \textit{Verification Cost}$$

Parsing Performance

Computational evaluation of PLTAG parser:

- train and test on standard Penn Treebank data (converted to PLTAG), with sentences of length 40 or less;
- assume gold-standard POS tags;
- use a supertagger to choose prediction trees (one word lookahead);
- coverage on the test set is not perfect: beam search; missing lexical entries.

Results

Model	Precision	Recall	F-score	Coverage
Baseline	44.39	52.38	48.06	85.10
PLTAG Parser	78.01	78.83	78.42	92.73
Prediction Tree Oracle	79.88	80.51	80.19	89.54

Baseline: pick most frequent tree (highest combined frequency of all subtrees).

Oracle: assume correct prediction tree (instead of supertagging).

Comparison with other TAG Parsers

Model	implemented	incremental	connected	predictive	F-score
Mazzei et al. (2007)	−	+	+	+	n/a
This work	**+**	**+**	**+**	**+**	**78.4**
Kato et al. (2004)	+	+	+	−	79.7
Sarkar (2001)	+	−	−	−	79.8
Chiang (2000)	+	−	−	−	86.7
Shen & Joshi (2005)	+	+	−	−	87.4*

*evaluated on dependencies

Comparison with other TAG Parsers

- Performance not directly comparable with parsers that use the original Treebank structure (simpler NP structure, etc.);
- there are structural differences even with other TAG parsers (LTAG, spinal TAG);
- Mazzei et al. (2007) parser conceptually most similar, but not implemented and evaluated;
- Kato et al. (2004) make strong simplifying assumptions (no modifier/argument distinction);
- Sarkar (2001) and Chiang (2000) parsers are not incremental;
- Shen & Joshi (2005) don't build connected structures.

Cognitive Plausibility

Psycholinguistic evaluation of PLTAG parser:

- train on Penn Treebank;
- take experimental materials from psycholinguistic experiments;
- parse them using the PLTAG parser, compute processing difficulty values for each sentence;
- compare to published reading time results.

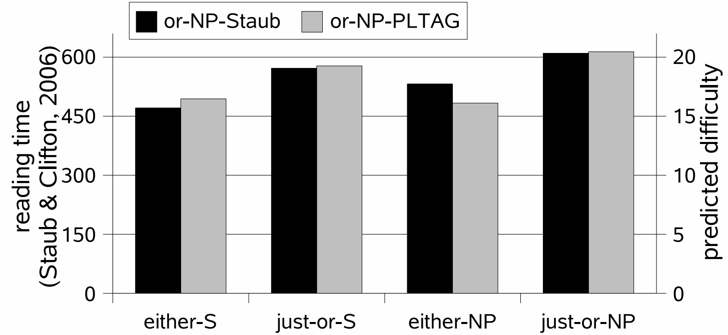
Baseline: standard surprisal model (PLTAG without prediction and verification component).

Either . . . *or* Constructions

PLTAG model predicts difficulty in *either* . . . *or* constructions:

(6) Peter read either a book or an essay in the school magazine.

(7) Peter read a book or an essay in the school magazine.

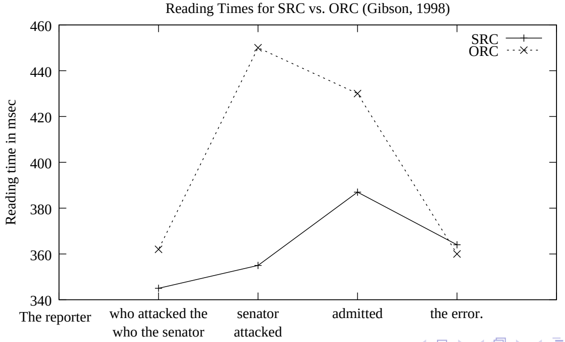


Relative Clause Asymmetry

Classic result in psycholinguistics: subject relative clauses are easier to process than object relative clauses.

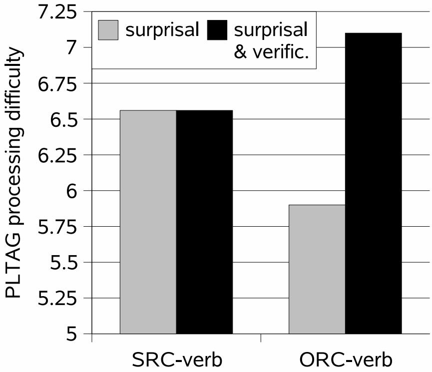
(8) SRC: The reporter that attacked the senator admitted the error.

(9) ORC: The reporter that the senator attacked admitted the error.



The relative clause asymmetry

PLTAG model predicts difficulty at verb region:



Correct predictions, but *verification component necessary,* results not predicted by surprisal-only baseline.

Conclusions

- Human sentence processing is incremental and predictive;
- evidence for lexical syntactic prediction (subcat frames);
- we presented a version of TAG that models these properties;
- the model comes with a parser, a probability model, and a linking theory;
- performance comparable to parsers with similar properties in the TAG literature;
- cognitive evaluation using experimental data: *either . . . or* prediction and relative clauses.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Chiang, D. (2000). Statistical parsing with an automatically-extracted tree adjoining grammar. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, (pp. 463–470), Hong Kong.

Demberg, V., & Keller, F. (2008). A psycholinguistically motivated version of TAG. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms*, (pp. 25–32), Tübingen.

Kato, Y., Matsubara, S., & Inagaki, Y. (2004). Stochastically evaluating the validity of partial parse trees in incremental parsing. In F. Keller, S. Clark, M. Crocker, & M. Steedman (eds.), *Proceedings of the ACL Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, (pp. 9–15), Barcelona.

Magerman, D. (1995). Statistical decision-tree models for parsing. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, (pp. 276–283), Cambridge, MA.

Mazzei, A., Lombardo, V., & Sturt, P. (2007). Dynamic TAG and lexical dependencies. *Research on Language and Computation*, *5*, 309–332.

Palmer, M., Gildea, D., & Kingsbury, P. (2003). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, *31*, 71–106.

Sarkar, A. (2001). Applying co-training methods to statistical parsing. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, (pp. 175–182), Pittsburgh, PA. Association for Computational Linguistics.

Shen, L., & Joshi, A. K. (2005). Incremental LTAG parsing. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, (pp. 811–818), Vancouver.

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either . . . or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425–436.

Sturt, P., Costa, F., Lombardo, V., & Frasconi, P. (2003). Learning first-pass structural attachment preferences with dynamic grammars and recursive neural networks. *Cognition*, *88*, 133–169.

Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, *29*, 291–305.

Vadas, D., & Curran, J. (2007). Adding noun phrase structure to the Penn Treebank. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, (pp. 240–247), Prague.