Sentence Comprehension as a Cognitive Process: A computational modeling approach
Day 1: An introduction to sentence comprehension

Shravan Vasishth and Felix Engelmann

Universität Potsdam, Germany
University of Manchester, UK
vasishth@uni-potsdam.de, felix.engelmann@manchester.ac.uk

August 15, 2016
What is sentence processing

Two central goals in this field are to understand
- online **parsing mechanisms** in human sentence comprehension
  - left-corner parsing, top-down, bottom-up? lookahead?
  - probabilistic parsing?
  - serial vs parallel vs ranked parallel?
  - deterministic vs non-deterministic parsing?
  - what kind of information is used to make parsing decisions (syntactic only, syntactic+semantic+...?)
- constraints on **dependency completion**
  - a general preference to attach co-dependents locally
  - constraints on retrieval processes
  - the consequences of probabilistic predictive parsing (expectation effects)
  - “good-enough” processing, underspecification, tracking only local n-grams (“local coherence”)
Introduction

- In this course, we will give a fairly narrow perspective on processing sentences out of context.
- We provide an extensive reading list on the course website for further details on the topics we mention.
- These slides also have references at the end.
- Please consult the references on the website and the ones cited in these slides for a fuller picture.
Introduction

Left-corner parsing, probabilistic parsing

Top–down

Bottom–up

Left–corner
Introduction: parsing mechanisms

Left-corner parsing [1], probabilistic parsing

**Left-Corner Parsing**

```
S → NP VP
N → man, dog
Det → a, the
V → ran, saw
VP → V NP
NP → Det N
```

**INPUT:** the
**GOAL CATEGORY STACK:** [ S ]
**ACTIONS:** If the is the left corner of any phrase structure rule then replace the stack content with the LHS of that rule. Repeat this left-corner rule until no further steps are possible. Wait for next input word. These actions yield the structure to the right:

```
S
   NP
     the
     N
   VP
```

**INPUT:** dog
**GOAL CATEGORY STACK:** [ N NP VP S ]
**ACTIONS:** Use the left-corner rule to expand dog to N. Since N is predicted in the incremental structure built so far (Step 1), integrate the N built up bottom-up into the tree. Since no further applications of the left-corner rule are possible, wait for the next input.

```
S
   NP
     the
     N
dog
   VP
```

**INPUT:** ran
**GOAL CATEGORY STACK:** [ VP S ]
**ACTIONS:** Use the left-corner rule to expand ran to V, and apply this rule once again to expand to VP. Since a VP is predicted in the structure, integrate this with the tree.

```
S
   NP
     the
     N
dog
r
   VP
```
Introduction

Left-corner parsing, probabilistic parsing

Purely top-down or purely bottom-up strategies turn out to be inappropriate models for human parsing [2, 3, 4] since they are unable to capture the observation [5, 468-470] that left-branching and right-branching structures are relatively easy to process compared to center embeddings:

(1) a. Bill’s book’s cover is dirty.
b. Bill has the book that has the cover that is dirty.
c. The rat the cat the dog chased killed ate the malt.
Introduction

Left-corner parsing, probabilistic parsing

More frequent attachments are preferred over rare attachments [6].

![Annotated Parse Trees for Two Interpretations of keep the dogs on the beach](image_url)
Introduction

Left-corner parsing, probabilistic parsing

Expectations for an upcoming verb phrase are sharpened if the verb’s appearance is delayed [7].
Introduction: parsing mechanisms
serial / parallel / ranked parallel

A general assumption in most work today is that parse choices are strictly serial. But theoretically, other options are possible, and there is some evidence for ranked parallelism [8].

- **Serial:** compute a single analysis, and if that fails, backtrack and compute new analysis (most classical theories, e.g., [9, 10, 11]).
- **Parallel:**
  - **Ranked:** Compute all analyses in parallel, but rank them (e.g., by likelihood).
  - **Prune:** using, e.g., beam search.
  - **Don’t prune at all:** generate all possible structures and then compute a function over them (e.g., entropy reduction, or surprisal) to find the optimal one [12, 13].
A common early assumption was that parsing was essentially deterministic.

A heuristic is to always prefer to attach locally [11]. Example:

(2) a. (low attachment)  
The car of the **driver** *that had the moustache* was pretty cool.

b. (high attachment)  
The **driver** of the car *that had the moustache* was pretty cool.

c. (globally ambiguous)  
The **son** of the **driver** *that had the moustache* was pretty cool.

Prediction: 2a,c easier to process than 2b.
[14] found/claimed that the word *moustache* was read fastest in the globally ambiguous sentence: the **ambiguity advantage**.
Introduction
deterministic / non-deterministic

One explanation [15] for this is to assume a non-deterministic race process (also see [16]):
Introduction: parsing mechanisms

Information sources: syntax only / all sources of information

[17] found evidence against syntax-first proposals, but [18] found evidence for syntax-first. (A too-common example of how prior beliefs of researchers are, uncannily, always magically confirmed.)

(3) a. The defendant examined by the lawyer turned out to be unreliable.

b. The evidence examined by the lawyer turned out to be unreliable.
Introduction: constraints on dependency completion

A local attachment preference

Non-local dependency completion tends to be more difficult than local dependency completion [19, 20].
Introduction: constraints on dependency completion

A local attachment preference

(4)  

a. The administrator who the nurse supervised scolded the medic while . . .  
b. The administrator who the nurse from the clinic supervised scolded the medic while . . .  
c. The administrator who the nurse who was from the clinic supervised scolded the medic while . . .
Introduction: constraints on dependency completion

A local attachment preference

Source: [20].
Introduction: constraints on dependency completion

Good-Enough processing / underspecification / local coherence

Source: [21]

(5) a. The coach smiled at the player who was tossed a frisbee
     b. The coach smiled at the player who was tossed a frisbee

Subjects seem to treat

(6) “the player tossed a frisbee”

as a main clause.
Introduction: constraints on dependency completion

Good-Enough processing / underspecification / local coherence
Introduction: constraints on dependency completion

Uncertainty increases with argument-verb distance (Safavi et al 2016)

(4) a. Strong predictability, short distance (PP)
   Ali a:rezouyee bara:ye man kard va…
   Ali wish-INDEF for 1.S do-PST and…
   ‘Ali made a wish for me and…’

b. Strong predictability, long distance (RC+PP)
   Ali a:rezouyee ke besya:r doost-dasht-am
   Ali wish-INDEF that a lot like-1.S-PST
   bara:ye man kard va…
   for 1.S do-PST and…
   ‘Ali made a wish that I liked a lot for me and…’

c. Weak predictability, short distance (PP)
   Ali shokola:ti bara:ye man xarid va…
   Ali chocolate-INDEF for 1.S buy-PST and…
   ‘Ali bought a chocolate for me and…’

d. Weak predictability, long distance (RC+PP)
   Ali shokola:ti ke besya:r doost-dasht-am
   Ali chocolate-INDEF that a lot like-1.S-PST
   bara:ye man xarid va…
   for 1.S buy-PST and…
   ‘Ali bought a chocolate that I liked a lot for me and…’

FIGURE 6 | The estimated entropy (with 95% confidence intervals), computed using the sentence completion data, for the two experiment designs.
Introduction: constraints on dependency completion

Uncertainty increases with argument-verb distance (Safavi et al 2016)

![Diagram showing reading times at the critical verb in Experiment 2.](image-url)
Introduction: constraints on dependency completion

Constraints on retrieval

Similarity-based interference has been implicated as a cause for difficulty in completing subject-verb dependencies. The essential idea is that retrieving an item (e.g., a noun) is harder (e.g., at a verb) if there are other competing items present that are similar on some dimension.

An implementation of this idea is Lewis and Vasishth (2005) (henceforth LV05), which is the subject of this course.
The model assumptions

This is often called “the” cue based model, but there are many cue-based models (Van Dyke’s, McElree’s conceptions are different from the LV05 model).

1. **Grammatical knowledge and left-corner parsing algorithm:**
   Note that a parser can do nothing without a grammar. So even asking a question like “is it the grammar or the parser?” technically doesn’t even mean anything.
   - If-then production rules drive structure building
   - Rules are hand-crafted in toy models, but scaling up has been done (Boston, Hale, Kliegl, Vasishth, Lang Cog Proc 2011).

2. **Constraints on memory processes affecting retrieval:**
   - allows us to model individual differences in attention and working memory capacity
   - Retrieval at any dependency completion point is a key (but not only) determinant of processing difficulty or facilitation.
### Introduction and background

The memory constraints in the model

<table>
<thead>
<tr>
<th>Latency factor F (:lf)</th>
<th>$RT = Fe^{-(f*A_i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$ Speed</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decay parameter d (:bll)</th>
<th>$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}) + \beta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$ Speed, forgetting</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source activation $W_k$ of buffer $k$ (e.g., goalbuffer :ga)</th>
<th>$A_i = B_i + S_i + P_i + \epsilon_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>This activation is distributed among goal-related chunks.</td>
<td></td>
</tr>
<tr>
<td>$\rightarrow$ Accuracy (goal-relevant), speed</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mismatch penalty $P$ (:mp)</th>
<th>$P_i = \sum_k PM_{ki}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$ Error sensitivity</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Similarity $M_{ki}$ between the value $k$ in the retrieval specification and the value in the corresponding slot of chunk $i$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$ Association between cue and target</td>
<td></td>
</tr>
</tbody>
</table>
Introduction and background

The memory constraints in the model: Similarity based interference
Introduction and background

The memory constraints in the model: Partial Matching

The tough soldier who Kathy met killed himself.

+ c-commander
+ masculine
+ c-commander
+ masculine

The tough soldier who Bill met killed himself.

+ c-commander
+ masculine
- c-commander
+ masculine
+ c-commander
+ masculine

* The tough girl who Kathy met killed himself.

+ c-commander
+ c-commander
+ masculine
Introduction and background

Possible evidence for partial matching: Processing polarity ([23] cf. [24, 25, 22])

Source: [22]

(7) a. No diplomats that a congressman would trust have ever supported a drone strike.
b. *The diplomats that no congressman could trust have ever supported a drone strike
c. *The diplomats that a congressman would trust have ever supported a drone strike.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7a) Accessible licensor</td>
<td>85</td>
<td>96</td>
</tr>
<tr>
<td>(7b) Inaccessible licensor</td>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>(7c) No licensor</td>
<td>83</td>
<td>86</td>
</tr>
</tbody>
</table>
Introduction: constraints on dependency completion

Constraints on retrieval

Consider again the Grodner and Gibson 05 results and our model [1] results:
Introduction: constraints on dependency completion

(8)  a.  *Target-match; distractor-mismatch*
    The surgeon\(^{+\text{masc}}\)\(^{+\text{c-com}}\) who treated Jennifer\(^{-\text{masc}}\)\(^{-\text{c-com}}\) had pricked himself\(\{\text{masc}\}^{\text{c-com}}\)\ldots

b.  *Target-match; distractor-match*
    The surgeon\(^{+\text{masc}}\)\(^{+\text{c-com}}\) who treated Jonathan\(^{+\text{masc}}\)\(^{-\text{c-com}}\) had pricked himself\(\{\text{masc}\}^{\text{c-com}}\)\ldots
Modeling retrieval processes in sentence comprehension

(9)  a.  Target-mismatch; distractor-mismatch
The surgeon $^{+fem}_{c-com}$ who treated Jonathan $^{-fem}_{c-com}$ had
pricked herself $^{fem}_{c-com}$. . .

b.  Target-mismatch; distractor-match
The surgeon $^{+fem}_{c-com}$ who treated Jennifer $^{+fem}_{c-com}$ had
pricked herself $^{fem}_{c-com}$. . .
Modeling retrieval processes in parsing
Modeling retrieval processes in parsing

Agreement attraction could also be an instance of similarity-based interference:

(10)  

a. The key_{sing} to the cabinet_{sing} is in the box.
b. The key_{sing} to the cabinets_{plur} is in the box.
c. * The key_{sing} to the cabinet_{sing} are in the box.
d. * The key_{sing} to the cabinets_{plur} are in the box.
Modeling retrieval processes in parsing

Lewis & Vasishth (2005) Parser

NP6
  cat : NP
  case : nom
  num : sing
  head : writer

NP14
  cat : NP
  case : acc
  num : plural
  head : editors

The writer who the editors from the journal supervised . . .

- Activation decay → distance effects
- Associative retrieval → Similarity-based interference
- Deterministic rule application → Expectation effects, reanalysis
Modeling retrieval processes in parsing
Engelmann, Jäger, Vasishth 2016


A probabilistic Earley parser as a psycholinguistic model.

The information conveyed by words in sentences.

Adjunct attachment is not a form of lexical ambiguity resolution.
Unrestricted race: A new model of syntactic ambiguity resolution. 

[16] Pavel Logačev and Shravan Vasishth. 
A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. 

Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. 


