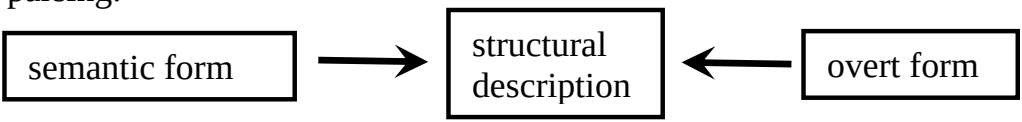
Lecture 4: OT Syntax

1. The nature of input in OT syntax
2. The generated outputs
3. Constraint inventory
4. Do-support
5. General discussion
6. Interpretive Parsing and how OT may overcome the competence-performance gap
7. Garden-path effects
8. Perception strategies and OT

0. Introduction: Core Ideas

- OT is not a theory of phonology proper but rather a theory of Grammar (and perhaps several other cognitive domains: semantics, vision, music.)
- The OT idea of robust (interpretive) parsing: competent speakers can often construct interpretations of utterances they simultaneously judge to be ungrammatical (notoriously difficult to explain within rule- or principle-based models of language)

- The presence of interpretable but ungrammatical sentences corresponds to mismatches between interpretive and productive parsing.



- The first part of this lecture outlines Grimshaw's OT account to grammaticality (including a factorial typology). This theory is founded on productive optimization.

- The second part explains interpretive parsing and introduces a constraint theory of processing. *Garden-path effects* of processing are predicted if optimal (interpretive) parses (corresponding to some early input) cannot be extended. This demonstrates that the principles of grammar have psychological reality for mature linguistic systems.

1 The nature of input in OT syntax

Following Grimshaw (1997), syntactic inputs are defined in terms of lexical heads and their argument structure:

INPUT

- lexical head plus its argument structure
- an assignment of lexical heads to its arguments
- a specification of the associated tense and semantically meaningful auxiliaries.

For convenience, we call such inputs *Predicate-Argument Structures* or simply *Logical Forms.*

Examples

- *What did Peter write?*

  {*write*(x,y), x=*Peter*, y=*what*, tense=past}

- *What will Peter write?*

  {*write*(x,y), x=*Peter*, y=*what*, tense=future, auxiliary=*will*}

Note that no semantically empty auxiliaries (*do*, *did*) are present in the input.

For treating embeddings more elaborated LFs are necessary (e.g. Legendre et al. 1998):

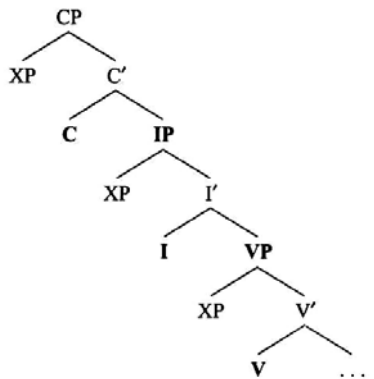
- *You wonder who eat what*
  wonder (*you*, Q_i Q_j *eat*(t_i , t_j))
  Q_i *wonder* (*you*, Q_j *eat*(t_i , t_j))

2 The GENerated Outputs

Minimal X' Theory

Each node must be a good projection of a lower node, if a lower one is present.

(X' Theory does not require that some *head* must be present in every projection!)



Extended Projection

An extended projection is a unit consisting of a lexical head and its projection plus all the functional projections erected over the lexical projection. The smallest verbal projection is VP, but IP and CP are both extended projections of V.

Example (continued)

$[_{VP}\,[_{V'}\,[_{V}$ write$][_{NP}$ what$]]]$,
$[_{IP}\,[_{NP}$ Peter$]\;\;[_{I'}\,[_{I}\,_\,]\,[_{VP}\,[_{V'}\,[_{V}$ write$][_{NP}$ what$]]]$
$[_{CP}\,[_{XP}\,\_\,]\,[_{C'}\,[_{C}\,\_\,]\,[_{IP}\,[_{NP}$ Peter$]\;\;[_{I'}\,[_{I}\,_\,]\,[_{VP}\,[_{V'}\,[_{V}$ write$][_{NP}$ what$]]]$

are all extended projections of $[_{V}$ write] (conform to further lexical specifications given in the input)

The GENerator (informal definition)

The core of GEN will construct all extended projections conform to the lexical specifications in the input. A further restriction is that no element be literally removed from the input ('containment'). The core can be extended by the following operations:

- introducing functional heads as they do not appear in the input, due to their lack of semantic content (e.g. the complementizer *that* and do-support in English
- introducing empty elements (traces, etc.), as well as their coindexations with other elements
- moving lexical elements.

Example (continued)

Input: {*write*(x,y), x=*Peter*, y=*what*, tense= past}

Some **Gen**erated outputs (using a simplified notation):

1. $[_{IP}$ Peter $[_{VP}$ wrote what]] ...*Chinese*
2. $[_{CP}$ what $[_{IP}$ Peter $[_{VP}$ wrote ***t***]]] ...*Czech, Polish*
3. $[_{CP}$ what wrote_i $[_{IP}$ Peter $[_{VP}$ e_i ***t***]]] ...*Dutch, German*
4. $[_{CP}$ what did_i $[_{IP}$ Peter e_i $[_{VP}$ write ***t***]]] ...*English*
5. $[_{CP}$ what $[_{IP}$ Peter did $[_{VP}$ write ***t***]]] ...*??*

Invalid outputs are

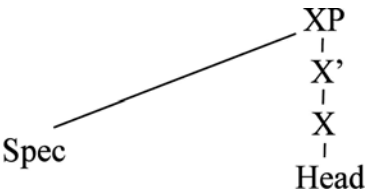
$[_{VP}$ wrote what]

$[_{IP}$ Peter $[_{VP}$ wrote _]]

$[_{CP}$ what $[_{IP}$ Peter $[_{VP}$ wrote what]]]

3 The constraint inventory

Markedness Constraints



- **Operator in Specifier** (OP-SPEC)
  Syntactic operators must be in specifier position
- **Obligatory Heads** (OB-HD)
  A projection has a head
- **Case Filter** (CASE)
  The Case of a Noun Phrase must be checked

Faithfulness Constraints

- **Economy of Movement** (STAY)
  Trace is not allowed

- **No Movement of a Lexical Head** (NO-LEX-MVT)
  A lexical head cannot move

- **Full Interpretation** (FULL-INT)
  Lexical conceptual structure is parsed
  (this kind of FAITH bans semantically empty auxiliaries)

OP-SPEC: triggers wh-movement wh_i ...t_i

OB-HD: triggers head-movement Aux_i ... e_i

4 Do-Support

> The auxiliary *do* is possible only when it is necessary' (Chomsky 1957)

Fact 1

Do is obligatory in simple interrogative sentences.

*What did Peter write? - *What Peter wrote?*

Fact 2

Do cannot occur with other auxiliary verbs in interrogatives.

*What will Peter write? - *What does Peter will write - *What will Peter do write?*

Fact 3

Do-support is impossible in positive declarative sentences.

*Peter wrote much - *Peter did write much*

Fact 4

The occurrence of auxiliary *do* is impossible in declarative sentences that already contain another auxiliary verbs, such as *will.*

*Peter will write much - *Peter will do write much - *Peter does will write much*

Fact 5

Auxiliary *do* cannot co-occur with itself, even in interrogatives.

*What did Peter write? - *What did Peter do write?*

The Analysis

- The auxiliary *do* is a semantically empty verb, one which only serves the *syntactic* function of head of extended projections.
- Do-support is triggered by the markedness constraint OB-HD at the expense of violations of the faithfulness constraint FULL-INT .

  OB-HD >> FULL-INT
- The facts of subject-auxiliary inversion in English suggest a ranking

  OP-SPEC, OB-HD >> STAY (see Exercice 2)
- Merging the two rankings

  OP-SPEC, OB-HD >> FULL-INT, STAY

  For English, the two markedness constraints outrank the general constraints (Faithfulness, Economy of Movement)

Example (concerning fact 1)

Input: {*write*(x,y), x=*Peter*, y=*what*, tense= past}		OP-SPEC	OB-HD	FULL-INT	STAY
1	[$_{IP}$ Peter [$_{VP}$ wrote what]]	*	*		
2	[$_{CP}$ what [$_{IP}$ Peter [$_{VP}$ wrote ***t***]]]		**		*
3	[$_{CP}$ what wrote$_i$ [$_{IP}$ Peter [$_{VP}$ e$_i$ ***t***]]]		*		**
4 ☞	[$_{CP}$ what did$_i$ [$_{IP}$ Peter e$_i$ [$_{VP}$ write ***t***]]]			*	**
5	[$_{CP}$ what [$_{IP}$ Peter did [$_{VP}$ write ***t***]]]		*	*	*

Fact 2 & 4: auxiliary=*will* in the input; same constraints & rankings.

Fact 3: Full Interpretation!

Fact 5: you have to assume that FULL-INT dominates STAY.

Typological consequences

In order to simplify discussion, the reranking approach to language typology ('factorial typology') will applied here to a very small set of syntactic constraints: {OP-SPEC, OB-HD , STAY}

- OP-SPEC, OB-HD >> STAY
  Both wh-movement and inversion occur in violation of STAY, to satisfy both top ranking constraints (example: *English*)
- STAY >> OP-SPEC, OB-HD
  Violations of STAY are avoided at the expanse of violations of 'well formedness'. A grammar arises lacking Wh-movement as well as inversion. (example: *Chinese*)
- OB-HD >> STAY >> OP-SPEC
  same picture as before

- OP-SPEC >> STAY >> OB-HD
  Wh-movement is forced but inversion cannot be used to fill the head position. A grammar arises that has Wh-movement but not inversion (example: *French*)
- Languages like *German* and *Dutch* require to consider the constraint NO-LEX-MVT (*No Movement of a Lexical Head*) which was undominated so far.
  Assuming NO-LEX-MVT to be outranked by the other constraints, structures like [$_{CP}$ Was schrieb$_i$ [$_{IP}$ Peter [$_{VP}$ e$_i$ ***t***]]] are optimal now (such languages are always incompatible with a semantically empty auxiliary).

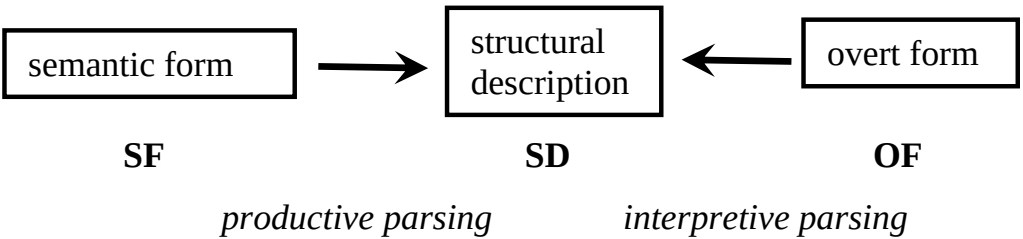
5 General discussion

- Bresnan (1998; see the reader) gives an important reformulation and improvement of Grimshaw (1995/1997; see the reader).
  - based on a mathematically sound structural account (feature structures in LFG)
  - adopts more radically non-derivational theory of *Gen*, based on a parallel correspondence theory of syntactic structures
  - conceptual and empirical advantages

- The problem of (language-particular) *ineffability*: There are input structures than can be realized in some languages but not others. For example, the questions “who ate what” is realizable in English and German, not in Italian. Such a question must be generable by *Gen* since it is realized in some language, and *Gen* is universal. Both in English and in Italian there is a non-empty candidate set. Consequently, in both cases there should exist an optimal output (a grammatical forms that expresses the question). But in Italian there is no grammatical form that means “who ate what”. (cf. Legendre, Smolensky & Wilson 1998)

6 Interpretive Parsing and how OT may overcome the competence-performance gap

- Human sentence parsing is an area in which optimality has always been assumed. According to the nature of (interpretive) parsing, in this case the comprehension perspective comes in: the parser optimises underlying structures with respect to overt form.



- Do the heuristic parsing strategies (assumed in the psycholinguistic literature) reflect the influence of the principles of grammar?

- Widespread and incorrect conviction that the impossibility of identifying the parser with the grammar had already been established with the failure of the 'Derivational Theory of Complexity' (e.g. Fodor, Bever, & Garrett 1974)

- Parsing preferences can be derived from the principles of UG if the proper grammatical theory is selected. There is evidence that in OT *the same system of constraints* is crucial for both productive parsing (OT syntax proper) and interpretive parsing. This finding is a first important step in overcoming the competence-performance gap. (See Fanselow et al. 1999)

7 Garden-path effects

Readers or listeners can be misled or ‘quoted up the garden path’ by locally ambiguous sentences

Example 1

- The boat floated down the river sank / and sank
- Bill knew John liked Maria / who liked Maria

Example 2

- While the cannibals ate missionaries drunk / they sang
- Since Jay always jogs a mile seems like a short distance / this seems like a short distance to him.

Garden-path model (Frazier 1979)

The parsing mechanism aims to structure sentences at the earliest opportunity, to minimise the load on working memory. In more detail:

- only one syntactical structure is initially considered for any sentence (ignoring prosody)
- meaning is not involved at all in the selection of the initial syntactical structure (*modular processing architecture*)
- the simplest syntactical structure is chosen (minimal attachment and late closure)
  - minimal attachment: the grammatical structure producing the fewest nodes or units is preferred
  - late closure: new words encountered in a sentence are attached to the current phrase or clause if this is grammatically permissible

8 Perception strategies and OT

Gibson & Broihier (1998) give a straightforward account how to implement the garden path model in OT. Following Frazier & Clifton (1996) a PSG is assumed in which there are no vacuous projections (generating, for example, $[_{NP}$ John] but not $[_{NP}$ $[_{N'}$ $[_{N}$ John]]]).

Inputs

Sequences of lexical items such as (*the, boat*) and (*the, boat, floated*).

Generated Outputs

The inputs are parsed into well-formed phrase structures (according to the rules of PSG). The actual output has to extend outputs of earlier inputs (in order to minimize the load on working memory)

(*the*) → output $_1$

(*the*, *boat*) → (output $_1$ + something) $_2$

(*the*, *boat*, *floated*) → (output $_2$ + something) $_3$

Constraints

- NODECONSERVATIVITY (correlate of Minimal Attachment)
  Don't create a phrase structure node
- NODELOCALITY (correlate of Late Closure)
  Attach inside the most local maximal projection
- NODECONSERVATIVITY >> NODELOCALITY

Garden-path effects are predicted if optimal parses (corresponding to some early input) cannot be extended.

Example 1 (continued) {*node conservativity* crucial}

1. (*the*)

   [NP [DET the]]

   (*Assuming the parser is top-down to some degree*)

2. (*the, boat*)

   [IP [NP [DET the] [N boat]]

3. (*the, boat, floated*)

   a. [IP [NP [DET the] [N boat]] [VP floated]]

   1 new node (VP) / 1 locality violation (NP)

   b. [IP [NP [DET the] [N' [N boat] [CP [IP [VP floated]]]]]]

   4 new nodes (VP, IP, CP, N') / 0 locality violations

Example 2 (continued) {locality crucial}

1. (*While, the, cannibals, ate*)

   [IP [CP [C while] [IP [NP the cannibals]] [VP ate]]]]

2. (*While, the, cannibals, ate, missionaries*)

   a. [IP [CP [C while] [IP [NP the cannibals]] [VP [V ate] [NP missis]]]]]

   2 new nodes (V, NP) / 0 locality violations

   b. [IP [CP [C while] [IP [NP the cannibals]] [VP ate]]]
   [IP [NP missis]]]]

   2 new nodes (IP, NP) / 3 locality violations (VP, IP, CP)

9 The constraint theory of processing (CTP)

The psychological reality of Grammar

Position A: Parser ≠ Grammar	Position B: Parser = Grammar
• early generativists • peoples shocked by the failure of the derivational theory of complexity (DTC)	• students following the DTC • some people believing in OT syntax (e.g. Pritchett 1992, Fanselow et al. 1999)

"Precompiled rules or templates are used in parsing" (Frazier & Clifton 1996). Such templates can be seen as a kind of procedural knowledge that gives an efficient, but rather indirect (non-transparent) realization of the grammar The psychological reality of grammatical principles is then at best confined to the role they play in *language acquisition*.	"If correct, this view argues against the necessity of specific assumption for design features of the parser - optimally, we need not assume much more than that the grammar is embedded into our cognitive system." (Fanselow et al. 1999) The principles of grammar have psychological reality for mature linguistic systems as well.

The basic idea of the **CTP** is that there is no difference between the constraints Grammars use and the constraints parsers use. “We may postulate that the parser's preferences reflect its attempt to maximally satisfy the grammatical principles in the incremental left-to-right analysis of a sentence.” (Fanselow et al. 1999: 3).

The following analyzes have an illustrating character only. We freely use abbreviations, e.g. *the boat* instead of *[$_{NP}$ [$_{DET}$ the] [$_{N}$ boat]]*. The symbols Comp, Infl indicate empty heads (with respect to CP and IP, respectively). OP_i indicates an empty operator.

Example 1 (again)

1. (*the, boat*)

   [$_{IP}$ the boat [$_{I'}$ Infl ...]

   1 violation of OB-HD)

   (*Assuming the parser is top-down to some degree*)

2. (*the, boat, floated*)

   a. [$_{IP}$ the boat [$_{I'}$ Infl [$_{VP}$ floated ...]

   1 violation of OB-HD)

   b. [$_{IP}$ the [$_{N'}$[$_{N}$boat] [$_{CP}$ OP$_i$ Comp [$_{IP}$ t$_i$ Infl [$_{VP}$ floated t$_i$]]]]] [$_{I'}$ Infl...]

   Many violations of OB-HD and STAY

Comments

- The first step illustrates *overparsing.* Postulating the IP-node and an (empty) Infl-Element we create a category that is able to check a case (satisfying CASE). The overparsing procedure can be seen as a way of finding a local optimum and is one of the key factors responsible for parsing preferences.
- In the second step there are two possibilities. Clearly, the option corresponding to "early closure" is preferred when evaluating the violations of the *grammatical* constraints.

Example 2 (again)

1. (*While, the, cannibals, ate*)

   $[_{IP}\ [_{CP}$ while Comp] $[_{IP}$ the cannibals $[_{I'}$ Infl $[_{VP}$ ate ...]

2. (*While, the, cannibals, ate, missionaries*)

   a. $[_{IP}\ [_{CP}$ while Comp] $[_{IP}$ the cannibals $[_{I'}$ Infl $[_{VP}$ ate missis ...]

   No new violations

   b. $[_{IP}\ [_{CP}$ while Comp] $[_{IP}$ the cannibals $[_{I'}$ Infl $[_{VP}$ ate]]]]

   $[_{IP}$ missis $[_{I'}$ Infl $[_{VP}$ ...]]]

   New violations of OB-HD etc.

Conclusions

The constraint theory of processing looks promising and is an opportunity to realize syntax as an psychological reality not only in the realm of language acquisition but also that of language comprehension. It is advantageous both for theoretical and empirical reasons

However, there are several questions:

- The precise foundation of overparsing.
- Are the constraints appropriate to derive *all* parsing preferences?
- The garden path effects are very different in strength. How to account for such differences in terms of OT?
- Extensions are required: the influence of world knowledge and prosody.