Universität Potsdam

Gisbert Fanselow, Matthias Schlesewsky, Damir Cavar, Reinhold Kliegl

Optimal parsing: syntactic parsing preferences and optimality theory

OPTIMAL PARSING:[*]

SYNTACTIC PARSING PREFERENCES

AND OPTIMALITY THEORY

Gisbert Fanselow, Matthias Schlesewsky,

Damir _avar & Reinhold Kliegl

IK Formal Models of Cognitive Complexity

University of Potsdam

INTRODUCTORY REMARKS

Principled accounts of syntactic parsing like the garden path theory (Frazier 1978, see Frazier & Clifton 1996 for further references) have always presupposed that the heuristic strategies that characterize the behavior of the human parser in e.g. the case of a local ambiguity are nothing but descriptive characterizations of more profound factors that come into play in parsing. Among the candidates for such factors are the limited capacity of the working memory (e.g., Frazier 1987), interference (e.g. Lewis 1993), the limited "window" size for the initial steps in parsing (e.g., Fodor 1998), or speed differences among competing analyses (e.g., Frazier & Fodor 1978). In the tuning approach (Cuetos & Mitchell 1985), on the other hand, parsing principles have an independent status of their own, and reflect the responsiveness of the parsing algorithm to frequency differences in the input.

The quite different view that heuristic parsing strategies reflect the influence of the principles of grammar (Pritchett 1992, Gorrell 1995, Phillips 1996) has received less attention and support in the past. Part of the reason for this may lie in the widespread yet incorrect conviction that the impossibility of identifying the parser with the grammar has been established in the seventies, with the failure of the 'Derivational Theory of Complexity' (see Fodor, Bever & Garrett 1974, Pritchett & Whitman 1995, Phillips 1996 for a discussion). Indeed, as we will briefly discuss below, most models of grammar cannot be applied directly in the context of left-to-right incremental parsing.

Similarly, Frazier & Clifton (1996:25) are certainly not alone in assuming that "precompiled rules or templates are used in parsing", that is, that *theorems* derivable from Universal Grammar are used in sentence processing rather than the axioms of UG themselves. The psychological reality of grammatical principles is then at best confined to the role they play in language acquisition. This position is far from being ruled out on *a priori* grounds - it is in fact the position originally taken in the generative approach, see Chomsky (1965).

We will argue in this paper that parsing preferences **can** in fact be derived from what Frazier & Clifton (1996) call the "transparent" application of the principles of UG if the proper grammatical theory is selected, viz. Optimality Theory (OT, Prince & Smolensky 1993), or similar models

that allow grammatical principles to be violated. 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. If correct, our position would also constitute a considerable step forward in the attempt to show that the principles of grammar have psychological reality for mature linguistic systems as well. Furthermore, our result would constitute a major argument in favor of specific proposals made in OT, because there is reason to believe that competing grammatical models like the Government-and-Binding approach (Chomsky 1981) or the Minimalist Program (Chomsky 1995, 1998) do not allow a similarly transparent application in parsing.

The paper is organized as follows. The first section outlines some of the basic assumptions of Optimality Theory, and argues that it is particularly suited for being applied in on line parsing. Section 2 discusses some general aspects of parsing with OT, presenting two different ways in which parsing preferences may arise. In section 3, we show how a number of major parsing preferences discussed in the literature can be derived from our approach. Section 4 is dedicated to the presentation and discussion of three experiments involving Case agreement effects, which we believe support the ideas presented here in a particularly strong way. Section 5. The paper concludes with two sections on parsing differences between language, and on the perception of ungrammaticality.

PRINCIPLE-BASED PARSING AND OPTIMALITY THEORY

In this section, we will discuss the following assumptions that constitute the heart of the OT-account of natural language, and that seem to render it particularly suitable for being integrated into a model of human sentence processing:

- OT grammars work with a set of universal principles. These are formulated in maximal generality, so they will often make conflicting predictions for individual constructions.
- Such conflicts are resolved by arranging the principles into language-specific hierarchies.
- If two principles make conflicting predictions, the principle with the higher rank wins.

Concrete proposals for OT grammars may spell out further details, which may sometimes be rather unhelpful for parsing discussions and should thus be modified - to the extent that this move does not go against the spirit of an OT grammar.

Parsing Preferences and Principle-Based Theories of Grammar

Only principle-based grammars like the Government and Binding Theory (Chomsky 1981) have a chance of making interesting predictions about parsing preferences. Consider the preference for interpreting locally ambiguous clause initial noun phrases such as *die Frau* in (1a) as subjects rather than as objects (see, e.g., Hemforth 1993).

(1) a. *die Frau   hat dem Kind   das Buch am Freitag gegeben*
          the woman has the-dat child the book   on Friday   given
          "the woman has given the book to the child on Friday"
      b. *dem Kind hat die Frau das Buch am Freitag gegeben*
      c. *das Buch hat die Frau dem Kind am Freitag gegeben*
      d. *am Freitag hat die Frau dem Kind das Buch gegeben*
      e. *gegeben hat die Frau dem Kind das Buch am Freitag*
      f. *das Buch am Freitag gegeben hat die Frau dem Kind*

Nearly all constituents can occupy the initial slot of a German sentence, as (1) illustrates. A construction-specific grammar has to list all these possibilities, for example in the form "a sentence may begin with a subject, followed by the finite auxiliary, in turn followed by ...." (=1a), or "a sentence may begin with an object, followed by the finite auxiliary, in turn followed by ..." (for 1b)). There is nothing *within* such a type of grammar that predicts why listeners apparently go for the former rather than the latter template when the morphology of the initial noun phrase does not decide between an object and a subject interpretation. Of course, we can add weights or preference

statements to the rules, or rank them, in order to capture the parsing preferences we find, but such statements would be add-ons to a grammar which does not use them in making predictions on what is a well-formed string.

The situation is different for modular, principle based grammars, as Pritchett (1992) or Gorrell (1995) observe. If the rules or constraints are in principle applicable to *all* syntactic structures (not just to particular instances), they have a chance of biasing the parser's decision into one direction at certain points. Optimality theory appears to be the only recent syntactic theory for which this holds without further manipulations of the basic theoretical approach.

A principle-based grammar like the Government-and-Binding-framework (Chomsky 1981) roughly has the following structure: there is a component that generates one or more structural representations for each sentence, and there is a principles/constraints component that specifies a set of conditions that must be met by these representations. In the GB-model, a structure is grammatical only if it satisfies *all* such principles. Consider now, for expository purposes, a specific principle like the -criterion (2):

(2)   -Criterion (non-standard formulation)
        Each argument expression (e.g. each noun phrase) must be linked to an argument place
    of a verb, and each argument place of a verb must be linked to an argument expression.

The -criterion captures the observation illustrated in (3) that sentences are bad if noun phrases like *the girl* in (3a) cannot be linked interpretively to the verb, or if one of the verb's argument position is left open, as in (3b).

(3)   a. *John arrived the girl
        b. *He told the girl

In parsing, we can observe a certain preference for the argument interpretation of a phrase in case of a local ambiguity, which can apparently be linked to (2). Thus, *that* could introduce a complement clause or a relative clause in (4), but there is a preference for the complement clause interpretation (exemplified in 4a) compared to the relative clause alternative (4b), see Frazier (1978), Altmann (1988).

(4)   he told the girl <u>that</u> .....
        a. the father has kissed the child
        b. the father has kissed the story

The initially prefered interpretation for the clause introduced by *that* is that it is an argument of the verb *tell.* This preference allows an *early satisfaction* of the -criterion: the requirements imposed by the thematic properties of *tell* are fulfilled when *that* is encountered and the parser postulates that *that* introduces a complement clause. On the other hand, the third argument of *tell* would still be missing if *that* was analyzed as introducing a relative clause modifying *tell*'s second argument. It is a natural move, then, to assume that the parser has the preference for the complement clause interpretation **because** "the -criterion attempts to be satisfied at every point during processing [...]" (Pritchett 1992:12).

More generally, 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. We sympathize with and subscribe to this view proposed e.g. by Pritchett (1992) and Gorrell (1995), but note first that there is at least a conceptual problem when one tries to apply the Government-and-Binding approach (or similar models) to parsing. This model assumes that structural representations must meet **all** requirements imposed by the grammatical principles in a **complete** fashion. Thus, (3b) is ungrammatical because it fails to meet the -criterion, it is ruled out by the theory of grammar - but note that (3b) corresponds to an early stage in the parsing of (4). We must suspend a too early application of the grammatical principles in order to not rule out (4) too early, but, on the other hand, they seem to have early effects. That the parser tries to satisfy the grammatical principles as early as possible does **not** follow from the grammatical concepts of

Government and Binding Theory, and the notion of a relative degree of fulfillment of grammatical principles is certainly alien to the GB-model.

Principle-based parsing with OT: A first example

The situation is different in Optimality Theory. OT shares with GB the assumption of a set of universal principles of syntax or phonology, and in many cases (most of those discussed below), even the individual principles which OT assumes look similar to what has been proposed in the GB-literature or in the Minimalist Program (Chomsky 1995). But OT concedes that these principles may be in conflict with each other, so that an individual structural representation cannot fulfill them all at the same time. For example, Universal Grammar requires in general that phrases should appear in their canonical positions (STAY!), but is also requires that operators should appear in positions that correspond to their semantic scope (PARSESCOPE). These principles run into conflict easily, as (5) shows. (5a) is in line with PARSESCOPE but fails to place *what* into the canonical object position - whereas *what* shows up exactly there in (5b), but appears in the complement clause although *what* must take scope over the complete matrix clause.

(5)   a.   what do you think that he bought
       b.   *you think that he bought what
(6)   Zhangsan   xiangxinshe   mai-le   shu
       Zhangsan   believe   who   bought   books
       "who does Zhangsan believe bought books?"

One of the two requirements has to be violated, then, and English and Chinese make different choices: English tolerates STAY violations in order to fulfill ParseScope, whereas Chinese does exactly the oppositive, as (6) illustrates. OT represents such constellations by assuming that the principles of UG are hierarchically ordered, such that a structure is grammatical if it has the best profile in terms of contraint violations, see (7).

(7)   A   is grammatical iff there is no   ' such that   and   ' compete with each other, and
       ' violates the highest principle   and   ' disagree on less often than   does.

In general, all grammatical structures violate some principles of Universal Grammar - a structure is grammatical if it violates principles of UG in a less drastic way than its competitors. This property implies that an OT grammar **can** be applied immediately to yet incomplete structural representations - that these violate grammatical principles is a property they share with complete grammatical structures. And, furthermore, the task one must carry out in online processing, viz. select the "optimal" possibility out of various alternatives, is exactly the one that is relevant for determining grammaticality in an OT grammar.

Let us apply this idea to a well-documented parsing preference. As mentioned above, there is a clear preference for the subject reading of a clause-initial noun phrase that is locally ambiguous between a subject and an object interpretation, as was established in Frazier & Flores d'Arcais (1989); for German see Hemforth (1993) for declarative clauses, Schriefers, Friederici & Kühne (1995) for relative clauses, and Schlesewsky (1996), Meng (1997) and Schlesewsky, Fanselow, Kliegl & Krems (in press) for wh-questions. Compare the representations (8a) and (8b) that might be built up and compete with each other for being pursued further when the parser has not done more than (a) hear or read *which woman* and analyzes it as a wh-noun phrase:

(8)   a.   subject interpretation
           [CP welche Frau [C' COMP [IP t [Infl' INFL [VP V]]]]
      b.   object interpretation
           [CP welche Frau [C' COMP [IP ? [Infl' INFL [VP V t]]]]

(8a) represents the subject reading of 'which woman', (8b) the object interpretation, with t standing for the trace of movement, as usual. (8a) and (8b) are 'bad' representations in the sense that they violate many principles - e.g., the verb position is not filled lexically. But if we evaluate these representations within an OT framework, this does not matter, because all well-formed grammatical structures violate grammatical requirements.   For the proper choice of the optimal

candidate, only the constraint violation profile matters. Since (8a) and (8b) do not differ with respect to, say, the lexical filling of the verb position (and no alternative structure could differ possibly, because the verb has not been in the input so far) this violation can simply be ignored. But observe, on the other hand, that (8a) fairs much better than (8b) with respect to quite a number of grammatical principles:

In (8a), the Extended Projection Principle (EPP, see e.g. Chomsky 1982) is respected in an obvious way, the subject position being filled by the trace of the wh-word. For (8b), it is less clear if the EPP is respected - if the subject position is complety empty, the EPP is violated, so (8a) does better on the EPP than (8b).

(9)   Extended Projection Principle (EPP)
          Every clause must have a subject!

Universal Grammar allows, however, for phonetically empty categories such as phonetically empty expletive (meaningless) pronouns, or phonetically empty argumental pronouns (as they occur in the subject position of Italian or Spanish, see e.g. Chomsky 1982). But if these are postulated in the subject position of (8b) in order to respect the EPP, violations of other grammatical principles arise. Thus, there is a principle of Full Interpretation (FI) argued for by Chomsky (1995) that requires that phrases make a contribution to the meaning of the clause -meaningless elements such as expletives violate FI on obvious grounds. Thus, if the subject slot in (8b) is filled by an empty counterpart of expletive *there*, (8b) would imply an FI violation that (8a) avoids. Similarly, if we postulate a phonetically empty referential pronoun in the subject position of (8b), we would arrive at a representation that is worse than (8a) on at least two grounds. First, empty referential pronouns need licensing (see Rizzi 1986), and the licensing element (say, explicit overt person-number inflection of the verb) is missing - a violation that (8a) avoids. Likewise, if there is a referential empty category in the subject position of (8b), it will incur a second violation of the   -criterion, because no verb was encountered so far that could provide an argument slot for the initial wh-phrase and this additional referential NP.

Furthermore, the movement from subject position to the clause initial slot in (8a) is shorter than the movement from object positon in (8b). Long movement is forbidden if a shorter one is possible (= the superiority effect, see Chomsky 1973, 1995, and Fanselow & Mahajan 1996, 1999 for a formulation relevant to our current problem). In (8a), the wh-phrase has been moved from the closest position possible. In (8b), the subject position was skipped - and unlike what holds for examples like (10) (movement of *you* would violate the requirement that constituent questions be introduced by a wh-phrase), nothing in the input licenses this violation of the principle that movement must come from the closest position possible - because, up to the current point in the parse, there is no input but the wh-phrase itself.

(10)   what did you see

Reflections on other principles of grammar might be added, but the general point will have become clear already: the partial structure (8a) is better than (8b) on a number of grammatical grounds, and it is not worse than (8b) in any other respect. Thus, (8a) will be selected if the parser follows the OT model because OT grammars always go for the optimal candidate (complete satisfaction of all principles is neither possible nor necessary) and the selection procedure does not differ at all from what happens in a standard OT grammar.

OT and Parsing: Some general but preliminary remarks

We may hypothesize, then, that an implementation of OT syntax principles in the parser allows to derive parsing preferences straightforwardly - and in an incremental fashion as new input becomes available. In the best of all possible worlds, **all** formally triggered parsing preferences reduce to grammar in this way.
The specific contribution OT can make to parsing does not only lie in the fact that its principles are violable, so that the conceptual problems do not arise that one may see with a direct application of standard GB theory to parsing. If grammatical principles may be violated when that

is required by more important constraints, they can take a much more general form than constraints that must always be surface-true in all languages. Therefore, violable principles have a higher chance of implying parsing preferences than inviolable ones. Compare (11) and (12) in this respect as possible formulations of the principle that implies the superiority effect.

(11)   cannot move to the clause initial position   of a wh-question if   is a wh-phrase
   and if there is a wh-phrase   that is closer to   than
(12)   cannot move to   if there is a   closer to   .

If we disregard so-called discourse linked wh-phrases (see Pesetsky 1987), the formulation in (11) for the superiority condition is surface true, at least for English, provided we have a proper definition of closesness:

(13)   a.   I wonder who you expected t to say what
        b.   *I wonder what you expected who to say t

But an application of (11) in parsing the initial segment (14) of a wh-question does not have any interesting consequences. It would just build up the (correct) expectation that *who* has not crossed a c-commanding (and not d-linked) wh-phrase on its path to the clause-initial position.

(14)   who .....

The situation is quite different with (12). If applied to (14), it implies the expectation that *who* has moved from the closed position compatible with the grammar of English, i..e, it implies the subject preference. But (12) can be a principle of English and German grammar only if it can be overridden, e.g. by a principle (the so-called wh-criterion of Rizzi 1991) that requires that the clause initial specifier position of a constituent question must be filled by a wh-phrase. Thus, although (12) favors (15a) over (15b), (15b) actually wins because (15a) does not respect the wh-criterion (provided the wh-criterion overrides (12)). (12) can take the general form it has (and that is very helpful in deriving parsing preferences) only because it is embedded in a context of other constraints many of which are more important than (12) in the case of a conflict. This point will be important in a number of further examples below.

(15)   a.   *it does not matter John t has invited who
        b.   it does not matter who John has invited

Note that OT claims that in case of a conflict between two principles, the one with the higher rank wins. Thus, although quite a number of considerations favor the subject reading of a clause initial wh-phrase, as we have just seen, only the constraint with the highest rank will actually decide between the subject- and the object interpretation. For most parsing preferences, this observation does not really play a role for the purposes of our paper, because the relevant grammatical principles seem to favor the very same candidate, but we will encounter at least one piece of evidence below that shows the effect of a ranking. The overall empirical predication OT parsing makes in this context is that the "strength" of a preference is not a function of the *number* of principles that favor it, in fact, the empirical prediction of OT parsing rather is that there is no such thing as the "strengt" of a preference (unless it is related to the rank of the decisive principle) - which seems correct if we don't confuse the "strength" of a preference to build up a certain structure with the relative ease or difficulty to undo initial decisions when they turn out to have lead the parser down the garden path, and if we confine our attention to the scope of the proposal, viz. grammatically triggered preferences.

OT also claims that the grammatical principles are *freely* rankable - at least from the perspective of grammar, since other consideration involving learnability, constraints on historical change etc. might imply that some rankings have too a low chance of realization only. At least for the parsing facts at hand, it is hard to see how this particular property of OT could be assessed empirically.

OVERPARSING AND LOCAL OPTIMIZATION

The claim that OT is a particularly adequate model for human parsing might seem quite surprising at first glance. In the OT models that are used in phonology and syntax, a GEN component generates a very large (potentially infinite) set of candidates that might be outputs (say, structural representations) corresponding to a given input (say, a set of words), out of which the EVAL component selects the optimal candidate. It seems to come close to a truism that first computing a large set of structural representations S, and then selecting one element of S, is much less effective than generating only one such candidate.

Parsing models that are not designed to be psychologically realistic (Maruayama 1990, Menzel 1997) may very well operate with a comparision of all structural possibilties for a given string and be quite successful in terms of e.g. robustness, but in general, optimality based parsing need not proceed in this way. As Tesar (1995:9) put it: "Although Optimality Theory is easily understood mathematically in terms of the generation and evaluation of all candidates in parallel, it is unnecessary, and in fact counterproductive, to consider computing optimal forms in those terms." Rather, what one needs is an effective incremental procedure for determining a small set of potential winners, in the best case of determining *the* optimal candidate incrementally. We will take up here and modify ideas put forward in Tesar (1995).

One of the central ideas in Tesar's work is the concept of "overparsing": nodes are postulated for elements that have not yet been perceived in the input. Thus - just as in certain implementations of principle-based parsing (see e.g. Crocker 1994) - one might postulate a full CP structure with empty slots for the complementizer, the Infl-node, the verb, etc. on the basis of a partial input consisting of a wh-phrase only - overparsing in the sense that more terminal nodes/heads are postulated than are necessary for the part of the input string at a given point in time. In such a model for parsing, we may follow Tesar (1995:46) in assuming that "overparsing operations may be repeatedly considered ... until none of them increases the harmony of the entries [in any of the cells]". To put it differently, overparsing operations applied to _ serve the purpose of constructing a grammatical representation with a constraint profile that is better than the one of _ , and they are repeated as long as overparsing improves the constraint violation profile.

Consider for example again the situation in which the initial segment of a clause that is to be processed is a wh-phrase like *who* or *which woman,* or their German, Italian or Dutch counterparts:

(16)   *welche   Frau*
        which   woman

It is immediately obvious that the human sentence processing mechanism must at least construct the representation (17a) for parsing this input

(17)   a.   [DP [D, wh *welche*] [NP *Frau*]]

(17a) is fine on a number of respects, but bad on others. It does not satisfy the principle WH-IN-SPEC (18) that plays a prominent role in the grammar of questions in an OT-framework (Müller 1997). (18) forces e.g. the movement of wh-phrases in so-called partial movement constructions like (19), but it can be violated if the specifier position is already filled as in German (20a) - but other languages like Bulgarian prefer to sacrifice the uniqueness of specifiers in the interest of (18), as (20b) shows.

(18)   WH-IN-SPEC
        A wh-phrase must be in the specifier position of a CP
(19)   was   denkst du   wen   sie   eingeladen hat
        what   think   you   who   she   invited   has
        "who do you think that she has invited"
(20)   a.   wer liebt wen
                  who loves whom
        b.   koj kogo mi_lis?
                  who what saw?

In any event, (17a) violates (18) on trivial grounds, and we can improve on that by overparsing the structure, that is, by postulating an (empty) Comp node[1], of which (17a) is the specifier, as in (17b). (17b) violates, however, a principle that (17a) respects: Obligatory Heads (see Grimshaw 1997), see (21).

(17)   b.   [$_{CP}$[$_{DP}$ [$_{D, wh}$ *welche*] [$_{NP}$ *Frau*]] COMP]

(21)   OBLHD
        The head position of a projection must be lexically filled

In fact, (21) roughly corresponds to the FILL Constraint that Tesar (1995) uses to restrict overparsing. Overparsing will take place only if (21) (or Fill) is outranked by a principle that can be satisfied by overparsing. Indeed, the constraat between (19) and (22) shows us, that WH-IN-SPEC outranks OLDHD (WH-IN-SPEC >> OLDHD): in (22), the COMP position of the lower clause is lexically filled by *that*, but *wen* was not fronted. That (22) is out, and (19) is in thus shows that WH-IN-SPEC is the more important principle.

(22)   *was denkst du dass sie wen eingeladen hat

If it is true (as seems to be the case) that (17a) and (17b) differ primarily in terms of (18) and (21), it is obvious that (17b) is the better representation for *welche Frau*. Would overparsing yet a further node lead to an even better representation? The answers seems positive, if we opt for INFL. By postulating this node, we create a category that is able to check a case for *welche Frau* - this is necessary by the principle (23) familiar since the early days of principle based grammars.

(17)   c.   [$_{CP}$[$_{DP}$ [$_{D, wh}$ *welche*] [$_{NP}$ *Frau*]] [COMP [INFL

(23)   CASEFILTER
        The Case of a Noun Phrase must be checked!

In fact, by postulating an Infl node, we are able to build up the representation (17d), in which the wh-phrase is linked to a trace we postulate in the specifier position of IP, satisfying the EPP that way, and the BIJECTION PRINCIPLE (Chomsky 1982 and much subsequent work) that requires that a wh-phrase binds a trace. (17d) is worse than (17b) on the ground that it incurs a second OBLHD violation (INFL is not lexically filled), but there is little reason to doubt that the CASE FILTER dominates OLDHD -- that the verb can fail to move to Infl in German, (Haider 1996) Dutch (Koopman 1996), or English (Pollock 1988), incurring an OblHd-violation thereby, indeed suggests that the presence of INFL in sentence structure when a subject must be licensed is due to a high-ranked requirement. Lacking evidence to the contrary, we postulate CASE FILTER >> OLDHD, by which assumption (17d) wins of (17b).

(17)   d.   [$_{CP}$[$_{DP}$ [$_{D, wh}$ *welche*] [$_{NP}$ *Frau*]] [COMP [$_{IP}$ t [INFL ...]]]]]

(24)   BIJECTION-PRINCIPLE
        There is a one-to-one correspondence between operators (e.g. wh-phrases) and
   variables (e.g. traces)

We may leave it open for the moment whether the further postulation of an empty V would still improve the situation. Postulating an empty V node would create a category Infl selects (viz., the VP), but it would incur a further OLDHD-violation. What is important is that going beyond the verb -say, by postulating an object position- would now worsen the contraint violation profile. To see this, compare (25a) with (25b)

(25)   a.   [$_{CP}$[$_{DP}$ [$_{D, wh}$ *welche*] [$_{NP}$ *Frau*]] [COMP [$_{IP}$ t [INFL [$_{VP}$ [V]]]]]]

---

[1] The postulation of other empty heads like, say, D or Infl would not do the job of improving on (18) for obvious reasons.

b.　　　　$[_{CP}[_{DP}[_{D, wh}\ welche]\ [_{NP}Frau]]\ [COMP\ [_{IP}\ \_\ [INFL\ [_{VP}[V\ \_\ ]]]]]]$

There seems to exist no constraint by which (25b) is better than (25a), but note, e.g., that by postulating a transitive verb construction we have arguably introduced two categories with a Case licensing potential, viz. Infl <u>and</u> the transitive verb (or the pertinent functional category that must obligatorily present in transitive structures, if your preferred theory of Case implies that). But at least since Chomsky (1993), Case assignment is assumed to be obligatority, in the sense that the general requirement for feature checking implies that feature checking/ assignment/licensing potentials must be made use of. Only one of the two Case checking potentials has been made use of in (25b), however. Similarly, if _ is the trace of *welche Frau*, we violate the Minimal Link Condition guaranteeing shortest moves, while it is respected in (25a). If _ is construed as the trace of the wh-phrase, (25b) involves a phonetically empty object that is not a trace, but such elements are also in principle banned from grammar unless licensed by specific elements not present in (25b), so (25b) is worse than (25a) on that ground as well.

Thus, the overparsing procedure roughly stops at (17d) - we do not improve our results when more empty heads are postulated. But (17d) is a representation in which the wh-phrase is a subject. We have derived the subject preference by

a. building up a tree representation which covers all terminal elements found in the input

b. "overparsing" the input, i.e. postulating empty heads and integrating them into the structure, as long as the result has a BETTER constraint violation profile than the structure without the head.

In general, we believe that the overparsing procedure as a way of finding a local optimum just outlined is one of the key factors responsible for parsing preferences.

Let us note in passing that the Active Filler Effects different from subject-object-asymmetries can be derived in our framework as well.

(26)　　a.　　　　what did she sing (t) about _
　　　　b.　　　　　　which patient did she bring t (the doctor)
　　　　c.　　　　　　which boy did she tell t Mary was incompetent
　　　　d.　　　　　　which boy did she tell Mary t was incompetent

The presence of *did* and *she* rule out a subject interpretation of *what* very early in (26a). Given that a verb phrase must be built up anyway, the most parsimonious way of fulfilling the Bijection principle for *what* is by inserting the trace into the direct object position of the verb. The dispreferred alternative structure in which _ is the trace of *what* does not differ from the preferred representation in the need of postulating a VP node, but it presupposes the postulation of a preposition head that is unfilled during the segment *did she sing*. Therefore, this alternative will not be considered initially. The same consideration applies to the (26c-d), with (26c) being the preferred option. The parser can avoid operparsing an empty Comp node necessary for (26d) because *which boy* could also be the object of tell, so overparsing Comp will not apply: the parser will work with the structure for (26c) when *she* rules out the subject interpretation of the wh-phrase. (26b) might be seen as more problematic[2], because the preferred option is in fact not grammatical – but note that the idea that *which patient* originated in the position immediately following *bring* is grammatical in the V NP PP construction, which is not ruled out by anything in (26b) when *she* excludes the subject reading for the wh-phrase. And for V NP PP, linking the trace to NP is preferred, because a link into PP implies again the postulation of an empty prepositional head which the parser avoids because of OBLHD unless a higher principle requires the violation of this constraint. Thus, the central Active Filler Strategy effects have been successfully derived in our model[3].

---

[2] Thanks to Lyn Frazier for pointing out this type of example to us.
[3] What we propose is vaguely related to approaches that compare chain lengths in the way proposed by Phillips (1995).

MOVEMENT AND REFERENCE SETS

The preceding section has shown that parsing preferences arise when local optimization arises from overparsing the input material, in a way that is constrained by the principle OBLHD, among others. Parsing preferences may arise independent of this consideration as well.

All current approaches to natural language grammar assume in one way of the other that movement is a costly operation. In OT, this has been captured as in (27), the principle STAY

(27)   STAY: *Trace

(27) forbids movement by penalizing the traces left over by movement. For a language like English, the wh-criterion is more important than STAY, so that the specifier position of a constituent question must be filled by moving a wh-phrase there.

Consider now (28):

(28)   the soldier saw the priest who was on the balcony

As Frazier & Clifton (1996:97) point out, there is a clear preference to interpret the relative clause as a modifier of the object in this type of construction, quite in contrast to the variable nature of attachment preferences one sees for relative clauses in other domains. This clear syntax based preference seems to arise exactly in those situations in which modification of one noun phrase (*the soldier*, in our case) presupposes an analysis in which the relative clause was extraposed, whereaus the other option does not necessarily involve this kind of movement. Clifton and Frazier thus propose that the observation concerning (28) follows from de Vincenzi's (1991) *Minimal Chain Principle* (MCP), that requires that no unnecessary members of syntactic cins be postulated, that is, which favors a non-movement analysis (a single chain member must be postulated) over a movement analysis (two chain members at least are necessary: the moved phrase and its trace). If the central difference between the two interpretations of (28) is indeed the question of the application of movement, then STAY favors the object modification analysis, in a sense that will be made clear immediately. In fact, the effects of the full MCP are either identical to the predictions of the Active Filler Strategy, or they follow from STAY – which means than MCP effects are derivable *in toto* from grammatical principles alone.

Let us assume, at least for the moment, that relative clauses are right adjoined to DPs, and may undergo an extraposition kind of movement, that is, let us ignore, at least for this moment, the alternative analysis proposed by Kayne (1994), Haider (1996), among others. Before the relative clause is encountered, a structure like (29) will have been built up by the parser:

(29)   [IP [DP the soldier] Infl [VP saw [DP the priest]]]

The integration of the relative clause signaled by the very next input element *who* implies some slight changes of the structure postulated so far. If it is to modifiy the object noun phrase, we have to tuck in a DP node (segment) between the DP node for *the priest* and VP, this new DP node is the new sister of *saw* and the mother of *the priest* and the relative clause. Similarly, the same operation would have to be applied to the subject noun phrase *the soldier* in case of a subject modification – but inserting there the trace of extraposing the relative clause rather than the relative clause itself, which, in the easiest case, will simply have been adoined to the highest IP node. In other words, syntactically, the two interpretations differ with respect to extraposition only. Given the principle STAY, the object modification analysis involving no extraposition will win.

With this example, we can highlight a very important aspect of the operation of grammar in online left-to-right-parsing: it is the lack of *a priori* knowledge about the lexical material and the meaning of the sentence to be parsed that leads to parsing preferences. This is a crucial aspect in deriving the preference in (29).

Preferences are predicted in the case of local ambiguities because of the (temporary) inaccessibility of constraining information. On obvious grounds, the algorithm computing grammaticality in OT presupposes a reasonable idea about which structures are in competition with each other. Thus, (30a) does not block (30b) although (30b) violates STAY, while (30a) arguably does not - or at least less often than (30b).

(30) a.   he came
     b.   what did you see

With a few exceptions (like expletive elements, pleonasitic verbs), one wants to say that S and T can be in competition only if they are built from the same lexical material - that they are based on the same "numeration", in more technical terms. Thus, (30a) cannot possibly block (30b), because their numeration sets differ.

In a straightfoward sense, then, the set of structures that are candidates gets smaller and smaller as we proceed in incremental parsing. This close to trivial property of parsing is responsible for the fact that parsing preferences do not translate themselves into statements on (un-)grammaticality directly. Thus, we predict a subject preference for *welche Frau* in (31), because the grammatical considerations that we apply in incremental parsing are constrained by the numeration set {*ich, weiss, welche, Frau*} only when the key decision is made, and the structure predicted is *küssten* is encountered, so that the structure we considered so far must be one compatible with the numeration set {*ich, weiss, welche, Frau, die, Kinder, gestern, küssten*}, which it is not, because the verb bears plural morphology, in contrast to the singular morphology of the alledged subject *welche Frau.*

(31) ich   weiss   welche Frau   die Kinder   gestern   küssten
      I   know   which   woman   the children   yesterday   kissed$_{pl}$

Thus, grammaticality is always based on the full numeration N, while parsing preference arise when subsets of N only constraint the possibilities of structure building.

But identity of numeration sets is a necessary, not a sufficient condition for two structures to compete with each other. Returning to our example (29), the readings in question have the same numeration – yet, the object modification analysis does not block the subject modification structure from the perspective of grammar – although it is the preferred analysis. Arguably, in the standard case, two structures compete with each other only if they have the same meaning (the same Logical Form (LF), in more technical terms, see Müller (1999). In slightly different terms, we might follow Legendre et al (1997) in assuming that a structural representation must try to be as faithful as possible to the input, that is, to the lexical material and some representation of the target meaning of the sentence. Given that subject and object modification of the relative clause imply different global interpretations for (29), the two alternatives do not compete with each other from the point of view of grammar, that is, from the point of view of a system that knows the whole numeration set and the global interpretation of the sentence.

But this is, certainly, not necessarily the perspective of the parser operating incrementally. The human parser does not try to assign a structural representation to a sentence in the light of a given interpretation – it must compute the interpretation simultaneously with the syntactic structure. Faithfulness to the final interpretation will, therefore, not be a central criterion for constraining competitions in parsing, for the simple reason that the interpretation of the sentence is, normally, not known beforehand.

We therefore propose that reference to lexical material and (intended) interpretation should be seen as factors potentially constraining the set of structures that compete with each other. From the omniscient perpective that god-like grammar takes, these factors will always extert their full force when the grammaticality of a given structure is considered. From the more modest perspective of the human parser, they can do so only to the extent they are known to a cognitive system operating incrementally. In other words, as long as the human parser does not have any a priori evidence concerning the semantic target of the relative clause in (29), the two structural options for

relative clause attachment compete with each other, because this competition is not blocked by interpretive facts set externally as unalterable.

The resulting syntax model comes much closer to the one proposed by Legendre et al (1997) than to the system envisaged by Müller (1999). Note also that our considerations allow quite nicely how contextual and other type of knowledge about likely interpretations of a sentence may influence parsing preferences: as soon as the parser *has* some quasi a priori knowledge that the relative clause modifies the subject, the structure pertinent to object modification does not fulfill the condition of being faithful to this kind of input information, and gets a low rank (Legendre-model) or leaves the competition (Müller-model).

Having made these necessary qualifications on which structures compete with each other, we can conclude this section by considering some more effects of STAY or the no-movement part of de Vincenzi's MCP. Recall that de Vincenzi (1991) suggests that the AFS should be reformulated as a *Minimal Chain Principle*: the parser prefers a non-movement analysis for a phrase P over an analysis involving movement if it has a choice (in this case, the chain of P has one element only), and it tries to keep the distance between chain links as short as possible (the AFS-part of the principle).

Empirical evidence for the descriptive validity of the Minimal Chain Principle comes from the observation that the object interpretation of NP (32b") is preferred over the subject interpretation (32b') for structures like (32a-b).

(32) a.   e V NP
      b.   ha   chiamato   il venditore
            has   called   the salesperson
      b'   "the salesperson has called"
      b".   he/she has called the salesperson

In Italian, the subject position of a finite verb may be left phonetically empty (the subject position is filled by the empty argumental pronoun *pro*, in syntactic terms), and subjects may undergo "free inversion", that is, they mave be moved to postverbal position (adjoined to VP, see e.g. Rizzi 1982 for this analysis). For verbs like *chiamare* "call" that allow a transitive and an intransitive interpretation, a sentence like (32b) is thus ambiguous: the subject may be realized as *pro* and correspond to no phonetic material all, in which case the overt noun phrase must be the object, or the overt noun phrase might be the subject of the clause, in which case it must have been moved to final position, with the subject position being filled by an invisible empty expletive.

As de Vincenzi (1991) shows, the latter option is dispreferred. This follows from the Minimal Chain Principle which tries to minimize the number of movement operations that must be postulated: if the subject is argumental and simply phonetically unrealized, and if the overt NP is the object, nothing has undergone (relevant) movement steps at all – quite in contrast to what holds if the subject has undergone an "inversion" movement. From a global perspective, the preference uncovered by de Vincenzi is also predicted if the "parser" tries to apply STAY. Under the subject interpretation of *il venditore,* this phrase has been moved from the position of e in (32a) to the position of NP - nothing of that sort happens when *il venditore* is the object.

It must be added, though, that the inversion analysis does not necessarily reflect the state of the art in the analysis of Italian. Since Koopman & Sportiche (1991), one would prefer analyses in which all subjects are base-generated (merged) in VP, either as a complement (unaccusative verbs) or as a specifier (unergative and transitive verbs). Thus, it is more likely that *il venditore* occupies a non-derived base position in (32) independent of whether it is a subject or an object. The two structures differ, then, more in terms of the phonetically empty category that occupies the preverbal position: it is a fully referential argumental empty *pro* for the interpretation (32b"), and an explitive meaningless empty pronoun for dispreferred (32b'). Consider Rizzi (1996) for a detailled analysis of these and other types of empty pronominals.

The preference for (32b") then seems to be related to the principle FULL INTERPRETATION (FI), which forbids the introduction of meaningless, expletive elements into a structural representation (see Chomsky 1993, Müller 1997). FI would exclude the expletive interpretation of *pro* when an argumental one is possible for obvious reasons. But note that FI normally conflicts with STAY, and does so for the structure under consideration as well: if subjects originate in VP, the assumption of argumental *pro* involves the movement of this category out of VP into clause-initial position, as indicated in (33b). Thus, although (33a) -but not (33b)- violates FI, (33b) violates STAY which, for *pro*, is respected in (33a). Given that both (34a) (violating STAY) and (34b) (violating FI) are grammatical, the easiest assumption would seem to be that STAY and FI are *tied* (Pesetsky 1997, 1998; Müller 1997) in Italian: the two principle have equal status, so that sacrificing one in the interest of the other is just as good as doing it the other way round. But if that is so, then neither FI nor STAY can really help us in establishing the parsing preference!

(33)   a.   [IP pro-exp [Infl Verb] [VP tverb DP]
        b.   [IP pro-arg [Infl Verb] [VP tverb tpro]
(34)   a.   Franca telefona t
        b   pro Telefona Franca

A closer look at (33a), the representation for the inverted subject analysis, reveals, however, why the preference does in fact arise. Empty expletives as we find them in inversion constructions do not exist in isolation, they need a DP ("the associate") they are coindexed with and with which they exchange features. So when the parser starts analysing initial *ha chiamato* for (32), it can arrive at (33b) as the representation of the initial hypothesis for (32b") very directly and without any overparsing operation (note it has already perceived the verb, and has postulated a subject which it can reconstruct into VP). Whether it can compute (33a) as a representation for (32b') is less clear: if it cannot, the structure [IP pro-exp [Infl Verb] [VP tverb] loses the competition with (33b) because the principle that requires one associate per expletive is violated (while everything else is *on par,*), if it can, (33a) loses to (33b) because the associate DP must in fact be projected from a D head inserted by overparsing which has not been encountered in the structure so far - consequently, (33a) violates OblHd at least one time more often than (33b). Given the new perspective on the actual structure of the two options for (29a), what seemed to by triggered by STAY turns out to be in fact due to OBLHD

STAY might also be expected to predict the fact the parser prefers a local movement analysis for phrases sitting in Spec,CP over an analysis in which this wh-phrase has undergone long movement out of a complement clause (see e.g. Fanselow, Kliegl & Schlesewsky 1999 for empirical evidence in line with the preference). This may be true under a global consideration of sentence structure, but in incremental parsing, OBLHD is, again, faster. As soon as the subject interpretation for, say, *who* in (35) is ruled out because of the input *you*, inserting the trace of *who* into VP is the cheapest way to respect the BIJECTION PRINCIPLE: for the long movement structure, a CP at least would have to be embedded into VP, implying the postulation of an empty Comp that the structure in which *who* is constructed as the main verb's object can avoid.

(35)   a.   who did you tell t that Mary came
        b.   who did you tell Mary t that Jane invited t

Thus, while STAY seems to be able to disfavor rightward movement analyses (as we have seen for relative clauses), leftward movement is more likely to be affected by OBLHD – STAY effects go into the same direction (so STAY will not undo the preference later), but they come in too late for being responsible for the first emergence of the preference.

Leftward movements which do not target the specifier position of a functional head –if they exist at all- may be an exception to this general statement. Thus, consider German scrambling as illustrated in (36) under this perspective.

(36) a. Gestern   küsste   die Frau   den   Mann
        yesterday kissed   the woman theacc   man
        "yesterday, the woman kissed the man"

b. Gestern   küsste   die Frau   der   Mann
    yesterday kissed   the woman the$_{nom}$ man
    "yesterday, the man kissed the woman"
c. dass   er   Tulpen   (gestern t)   wässerte
    that   he   tulips   (yesterday)   watered

(36a-b) show that the subject (36a) and the object (36b) may be the first noun phrase following the finite verb in a main clause. The examples chosen here involve a local ambiguity for this first noun phrase, which is resolved by the case marking of the second noun phrase. There is ample evidence that these data show a subject preference for the first noun phrase, too.

According to the standard analysis of word order variation within the clause in German (see, e. g. Deprez 1989, Fanselow 1988, 1990, Mahajan 1990, Webelhuth 1987, and the contributions in Grewendorf & Sternefeld 1990), the underlying order of German is subect before object, with object before subject order being derived by an application of movement, viz. scrambling. Thus (36a) has an abstract structure such as (37a), and (36b) an abstract structure like (37b).

(37)   a.   XP verb[$_{IP}$ subject [$_{VP}$ object ....]
       b.   XP verb[$_{IP}$ object [$_{IP}$ subject [$_{VP}$ trace(object) ....]

When the first noun phrase is encountered in parsing (36a-b) incrementally, the verb has already been analyzed, so the postulation of the IP and the VP in (37a-b) cannot induce any OBLHD violations. In fact, one may argue that the reason why *die Frau* is preferentially analyzed as a subject in (36a-b) is the need to assume an application of movement for an IP-initial object that we can avoid for subject-initial structures. Consequently, we might attribute the preference we observe to STAY, but note than an object analysis of *die Frau* in the input segment *gestern küsste die Frau* induces an EPP violation as well, or an FI violation if we try to repair the EPP violation by inserting an empty expletive. Given the little evidence we have concerning the ranking of the EPP and STAY in German[4], we cannot be sure that the preference is due to one principle and not to the other.

The situation is different with examples such as (36c). Objects come fairly late in the canonical serialization of a German clause (see e.g. Lenerz 1977), so one would assume that *Tulpen* is scrambled to the pre-adverb position in (36c). Since the subject has already been filled in (36c) before the object is encountered, any effect concerning the preferred position of the object relative to, say, adverbs, cannot be due to EPP effects on trivial grounds. In fact, if the scrambling analysis is correct, STAY the primary and perhaps only principle that implies the expectation of a late object, or, to put it differently, that implies the expectation that nothing but elements from a certain class of verb-related should be able to follow the object. More generally, we would expect that clauses that show canonical order in the VP should be easier to read (because no reanalysis is necessary) than clauses with scrambled orders. Indeed, in a simple self paced reading time experiment, we found the expected effects for verb phrases structures as in (38), i.e. involving one or two objects and a PP: disrespect for canonical order implies longer reading times.
(38)   a.   NP PP V
       b.   NP NP PP V

The scrambling effect thus seems to be a consequence of STAY, at least as long as we assume scrambling involves adjunction and not movement to specifier positions of additional functional projections, in which case OBLHD as applied to the heads of these might outrank the STAY effect without contradicting it, though. Finally, Farke (1994) reports a subject preference for the initial NP1 in infinitival complements of matrix verbs like *lassen* "make, have" in German – this could be a STAY effect, but other factors might come into play as well.

(39)   Subject [NP1 NP2 V] lassen "make"

---

[4] If German has no empty expletives, the fact that the subject can stay in VP in German shows that Stay >> EPP in German, however.

FURTHER PARSING PREFERENCES

In this section, we will consider some further parsing preferences discussed in the literature, in order to see if they can be derived from an on-line application of grammatical principles as well.

We have seen that the on-line application of OBLHD makes a number of excellent predictions for sentence analysis, because it seems to constrain overparsing in just the right way. In this respect, OBLHD is, of course, not much different from the quite similar approach of Gorrell (1995) who tries to ban superfluous structure building as such.

Consider now the classical garden path effect in (40) discussed first by Bever (1970):

(40)   the boat floated down the river sank

When clause-initial *the boat* is encountered, overparsing involving at least an Infl node is likely – by this move, the DP *this boat* gets the required Case. If *floated* is integrated into the structure as the main verb, as in (42a), nothing spectacular happens in terms of constraints. For the reduced relative clause analysis as in (42b), we not only need to tuck in a DP node between IP and the DP dominating *the boat* (which may be tolerable) - building up this extra structure does not improve on any of the principles of grammar as compared to (42a), but it is worse on many other grounds: the empty relative operator we need to postulate induces at least one STAY violation, and the Comp and Infl nodes needed in the reduced relative clause induce OBLHD violations. (42a) being the competitor, (42b) simply has no chance to win.

(41)   [IP the boat [Infl ...]

(42)   a.   [IP the boat [Infl [VP floated ...]
        b.   [IP [DP [DP the boat] [CP OPi [Comp [ti Infl [VP floated ti ...]]]]]] Infl ...]

Note in particular that the postulation of an empty Comp heading a relative clause attached to *the boat* (the first overparsing operation for (42b) makes the constraint violation profile worse rather than better, as compared to what holds for *the boat* alone. Alternatively, we might suspect that *floated* itself is overparsed as a CP before it is attached to the NP in the structure (42b). Under this perspective, the verb –that is, the fact that it will not be overparsed as a CP in the absence of compelling evidence- is made responsible for the garden path effect. Intuitively, this is way of approaching (40) seems more correct, but we have no empirical evidence that would bear on this issue.

Consider now the following versions of the classical example (40) as discussed by Pritchett (1992). On the one hand, these examples seem interesting because they imply certain complications for a number of conceivable accounts of the core effect, as Pritchett notes. Thus, approaches that claim that the matrix verb reading is favored over the reduced relative analysis on the ground that the matrix verb allows to assign Case to the clause initial DP fail to predict the disadvantage for reduced relative clauses in (43), where *believe* and *know* would be able to provide Case for a DP *the horse*+reduced relative clause.

(43)   a.   John knows the horse raced past the barn fell
        b.   John believes the horse raced past the barn fell

Similarly, the fact that *float* can theta-mark the DP *the boat* if it is a matrix verb, while the larger DP *the boat* + free relative formed by *floated* would still be without a thematic role when *floated* is processed is not likely to be the best or only reason for the preference, because *knows* can in principle theta-mark an NP beginning with *the horse*, independent of whether it is modified by a reduced relative clause or not.

For our model, (43a) poses no problem, although the preferred analysis for *John knows the horse* ... is one in which *the horse* is a verbal object and not the subject of a complementizerless clause (see also below). Thus, at the point when *raced* is encountered, the parser will have postulated (44)

(44)   [IP John Infl [VP knows [DP the horse]]]

The next item that is processed, viz. *raced* cannot be integrated into the parse tree without *the horse* being de-linked from VP first in both potential analyses. This is obvious for the (preferred) clausal subject interpretation and holds for the reduced relative clause reading as well, because the relative clause will be adjoined to DP rather than tucked into it. In particular if the view is correct that the failure of *raced* to be overparsed is the cause for the garden path effect, the reduced relative interpretation again has no chance of being built up.

There is a class of structural preferences typically labelled "late closure" effects which we established in detail by Frazier (1978) and which follow from an application of OBLHD in a straightforward way. Thus, consider the structures in (45), in which the "unprimed" structures correspond to the preferred structural alternative:

(45)   Object -subject ambiguities
         a.   she found out [DP the answer to the physics problem] quickly
         a'.   she found out [IP [DP the answer to the physics problem] Infl was easy]
         b.   while Mary was mending [the socks] she fell asleep
         b'.   while Mary was mending [IP [DP the socks] Infl fell off her lap]
         Complement/subject ambiguities
         c.   [PP in this race to touch the wire] she will win
         c'.   [PP in this race] [IP [IP to touch the wire] Infl is to die
         d.   [PP without her stupid remarks] this paper would have been much better
         d'.   [PP without her][IP [DP stupid remarks] Infl would not have been eliminated
         from the paper

In all examples in (45), there is a local ambiguity for the underlined segment, that could either be the complement of a preceding head, of the subject of a following clause. Obviously, the subject interpretation implies the overparsing of at least an Infl node which violates OBLHD, without this being justified by anything that can be found in the material parsed up to the relevant point. Thus, garden paths are correctly predicted to arise in examples(45a`- d'). This analysis works smoothly in (45) in particular if we assume that *her* is a D-head that either takes a complement (possessive interpretation) or not (full pronoun).

Note, furthermore, that the parsing behavior of the clause-initial PPs in (45c-d') differs crucially from clause initial noun phrases. Noun phrases need Case, and this need for case drives the overparsing of an Infl node, so that IP is constructed on the basis of a clause initial NP. PPs need no case, so there is no reasing for postulating an IP node by overparsing during the processing of the PPs above. If matters were different, the overparsed IP would offer a slot for a subject, and the EPP would force *stupid remarks* and *to touch the wire*, respectively, into the subject position, implying incorrect parsing predictions thereby. The lack of overparsing with PP (it would not satisfy any need of PP, and must therefore be avoided given OBLHD) is thus crucial.

The examples in (46) and (47) discussed by Philipps (1996:114-130) and Gibson & Broihier (1998. 171-172), respectively are particulary interesting. At least in (46a), a preference for attaching *I made* low as a relative clause (as in *because Rosa praised the recipe I made, I sent her a copy of it)* seems to have been established experimentally, but the temporal examples showed a preference for the matrix subject-verb interpretation (as for; *after Mary got off the bus, she bought a candy,* see the experiment described in Philipps, loc. cit) – in contrast to the intuitions reported by Gibson and Broihier for temporal (47).

(46)   a.   Because Rosa praised the recipe I made ...
         b.   After Mary got off the bus she bought ...

(47)   while I talked with the woman John was ignoring ....

Notice that the relative clause interpretation involves the overparsing of a Comp, an Infl and a verbal head, and a STAY violation of an empty operator:

(48)   praised [the recipe [_CP_ OP [Comp [_IP_ I Infl [_VP_ V t]]]

In contrast, the subject reading for *I* involves a single OblHd violation in the optimal case for this interpretation. Thus, preferences for (46b) but not for (46a) are correctly predicted by our model. At present, we cannot account for this difference.

(48)   [because Rosa praised the recipe] [_IP_ I Infl

Speakers of English prefer the complement clause introducing interpretation (49a) of *that* in (49) over the one in which *that* is the first item in a relative clause (49b). To the extent that relative clauses form part of the noun phrase they are semantically linked to, this goes counter a late closure strategy, but OBLHD favors neither of (49a,b), since both structural possibilities have a filled Comp. But recall that the relative clause interpretation requires the postulation of an empty relative operator OP preceding *that.* This OP element does not appear in the complement clause interpretation, and incurs a violation of STAY that is absent in (49a).

(49)   John told the girl that
          a.   we invented the story
          b.   we kissed the story

The preference for the object interpretation of *the dog* in (50) over an analysis in which it is the subject of a relative clause (in parentheses) may be linked to either the three additional OBLHD-violations (the relative clause involved an empty Comp, and since *the dog* is the subject, an Infl node must be postulated, too, finally, the empty operator then cannot help but being linked to a trace lower than the subject position, i.e. in VP, for which we need an empty verb), or to the STAY-violation by the empty operator

(50)   John gave the man the dog (bit a package)

Parsing theory had to cope with the fact that it needed to assume two heuristic principles, viz., Late Closure and Minimal Attachment, that contradict each other in many cases. The empirical findings of Konieczny et al. (1997) establish, however, that the situation is more complex than a simple conflict of two different parsing strategies/laws of grammar. Consider first (51) in this respect:

(51)   a.   dass er den Kuchen   mit den Kirschen ...
                that he the cake   with the cherries
        b.   dass der Arzt   der Schauspielerin
                that the doctor   the actress

(51a-b) constitute initial segments of *verbfinal* clauses in German. The PP following *den Kuchen* "the cake" in (51a) could either be attached to this object noun phrase, or be attached to a VP the head of which has not yet been parsed. Similarly, the NP *der Schauspielerin* "the actress" in (51b) could either be a genetive attribute to "the doctor", or the dative object of a verb yet to be heard. Empirical evidence from eye tracking studies suggests that the parser disprefers VP-attachment for PP and NP in this case.

These observations can be captured easily if we are more explicit on a particular aspect of overparsing. In (51b), *der Arzt* is recognized as a subject on the basis of its morphology, and for the assignment of nominative case, we need to overparse an Infl node. If the structure postulated on the basis of the first two words following the complementizer *dass* in (51b) is thus just (52), we can understand why *der Schauspielerin* preferentially attaches to *der Arzt*: this analysis requires no new nodes, while a VP-attachment implies the assumption of an empty V violating OBLHD on obvious grounds.

(52)   [IP [DP der Arzt] Infl

Konieczny et al. demonstrate furthermore that parsing preferences are reversed in verb-second structures such as (53). Now, the parser seems to follows the predictions of Minimal Attachment:

(53)   a.   er verzierte   den Kuchen   mit den Kirschen ...
             he decorated   the cake   with the cherries
         b.   wahrscheinlich   gefiel   der Arzt   der Schauspielerin ...
             probably   pleased the doctor   the actress

Notice that the presence of *gefiel* in the input material preceding *der Schauspielerin* necessarily implies the prior postulation of a verb phrase of which *gefliel* is the head – together with its feature checking dative case. This case feature can be checked by *der Schauspielerin* if this noun phrase is an object. On the other hand, the genetive case that shows up in noun phrases is not a result of a lexically specified property of nominal heads, rather, it seems to be an optional feature arising in noun phrases. Thus, we do not leave any lexical feature unchecked when *der Schauspielerin* is NOT attached into NP - but doing the latter would leave the lexical case feature of *gefiel* unchecked, so NP attachment is dispreferred as soon as the verb has been encountered. The difference in attachment preference between (51b) and (53b) is therefore derived, provided that we do not insert an empty V node into (52) immediately. In fact, it is not at all clear whether there is a principle that would override the OblHd violation incurred by an empty V-head. In the absence of evidence to the contrary, we thus take our account to be correct.

What about the PPs, then? According to Frazier (1998), the empirical evidence concerning (54) certainly disconfirms the idea that there is a structural preference for attaching the PP to NP, but that there is a clear structural preference for VP attachment independent of the choice of the verb seems also to not have been firmly established.

(54)   he saw the man with the binoculars

Pritchett (1992: 146-148) accounts for a VP attachment preference by assuming that PPs are quasi-arguments of the verb. In fact, Alexiadou (1994, 1997) and Cinque (1999) argue for the linking of at least certain adverbs and adverbial PPs to functional heads in sentential structure – the visible projections would check corresponding features with these heads. Note that the presence of the verb renders the postulation of such heads virtually cost-free in English and German, namely if we assume (certainly correct for English) that the verb moves through these functional heads, so that they are not empty in the sense of OBLHD – they contain the trace of the verb. The additional functional head for the "VP"-attachment of the PP will imply a further STAY violation by the verb (the verb moved through this head), but this STAY violation either arises in the noun phrase as well, or is outweighed by the advantage of checking an interpretive feature of the PP by the functional head in VP. Since there is no evidence for corresponding functional heads in (underived) noun phrases, the VP attachment preference thus reduces to the assumption that the checking of a PP feature is advantageous – but never more important than OBLHD. Otherwise, we would not only fail to account for the NP attachment preference we see in (51a), but we would also lose the account of (45c-d).

Our discussion has so far been deliberately been silent about two aspects, one principle and one set of data, that are usually highlighted when parsing and the relation between the grammar and the parser are topics.

The principle we have been silent about in the last sections is the theta-criterion (55), although it played the major role in Pritchett's (1992) attempt of deriving the parser from the grammar. On the one hand, we have seen that reference to (55) is not needed to account for the parsing preferences we observe, and it may make incorrect predictions in certain domains (see Frazier & Clifton 1996:22) that do not arise when formal feature checking in the sense just outlined is invoked.

(55) _-Criterion (non-standard formulation)

Each argument expression (e.g. each noun phrase) must be linked to an argument place of a verb, and each argument place of a verb must be linked to an argument expression.

On the other hand, the _-Criterion should not figure in our considerations at all, because it is not a violable principle. In early accounts of non-configurational languages such as Japanese or Warlpiri like Hale (1983), the theta-criterion was not assumed to be surface-true for all languages, but more recent treatments of the phenomenon such as Baker (1996) and Fanselow (submitted) avoid such a move. Furthermore, there is no clear example of a construction of English or German in which the _-Criterion would be sacrificed in the interest of some other principle. The _-Criterion is not among the competing principles.

We would certainly run into a conceptual problem if the only option we have would be to assume that (55) is among the principles that govern the generation of potential candidates for the competition, and which are thereby inviolable. We cannot avoid that initial segments of parses violate (55) partially, and the handling of this problem would then lead us outside grammar, again. But note that there is not too much evidence (see Chomsky 1995 for some remarks, and Fanselow, submitted, for a criticism) that (55) is indeed part of the *grammar* of natural language. It makes as much sense to follow Chomsky's (1965:157ff.) orginial intuition that the well-formedness of (57) implies that disrespect for (55) is more a matter of creating "gibberish" in the sense of Chomsky (1995) than generating a formally incorrect statement. Note also we seem to have evidence (see McElree & Griffith 1995) that thematic role information is used later than formal syntactic subcategorizatiion information in online parsing, and should therefore not figure in formulating initial preferences at all.

(57)   it is nonsense to speak of a man arriving a cat

On the empirical side, note that certain late closure effects cannot be made follow from OBLHD or similar principles of grammar.

(58)   I met the man who will kiss her yesterday

Empirically, we observe a preference for a late closure analysis of *yesterday* rendering the sentence (58) difficult to parse. From the perspective of OBLHD, neither low nor high attachment of *yesterday* is favored over the other. *Ceteris paribus*, grammatical principles do not care about the place at which a phrase is attached to, as long as overall requirements are met in the same way.

Pritchett (1992:113) notes that some speakers are garden-pathed by (59a), while the same is true with (59b) for others. If this is correct, the expectations derivable from our model would indeed be borne out: for the satisfaction of grammatical principles like checking the case of *the fudge*, it does not really matter which category helps to satisfy the principle. Unfortunately, Pritchett's claims are based on informal studies, only.

(59)   a.   Katrina gave the man who was eating the fudge
        b.   Katrina gave the man who was eating the fudge the wine.

The preference exemplified in (58) seems uncontested, however, although it need not necessarily "attachment" in the strict sense, if one takes up the empirical observations and theoretical proposal made in Frazier & Clifton (1996).

Instead of speculating, however, whether there is a merely interpretive preference to link an adjunct to the thematic domain that is currently processed, we may also observe that there may be very "shallow" ordering principles for English that require, e.g., that finite clauses should appear at the periphery of syntactic domains. Due to this (violable) ordering principle, the parser will build up a structure in which the relative clause is marked as right peripheral in (58), and this right-peripherality is certainly incompatible with the appearance of yet another VP consituent, viz. *yesterday*. While they may be correct, these considerations have little to say about Abney's (1989) (59), showing a clear preference for interpreting the box as the boy's location.

(59)   a gift to a boy in a box

Fodor (1998) discusses the idea of relating late closure effects like the problematic one we discuss right now to the packaging idea of the original "sausage machine model" of Frazier & Fodor (1978). Given that the observable attachment preference for the adjective *divorced* varies in (60) as indicated there, Fodor suggests that such preferences are a consequence of a preference for syntactic boundaries (sketch by underling/italicization) and phonological boundaries to coincide, with an additional preference for "phonological sisters" to have roughly the same size. Thus, the attachment preference for *divorced* in (60) seems to be a function of the size of the resulting syntactic and phonological packages.

(60)   a.   the divorced bishop'*s daughter*   (bishop)
        b.   the recently divorced *bishop's daugher* (daughter)
        c.   the recently divorced bishop's *daughter in law* (bishop)

Returning to (58), it is easy to observe that we get a more equal distribution of weight if the clause is phrased as *I met the man* plus *who will kiss her yesterday,* than in the case if a high attachment of *yesterday*, which would have to be realized involving the intonational units *I met the man, who will kiss her, yesterday*.

If this account is true and is able to capture most of the remaining late closure effects, it has a chance of constituting considerable additional evidence for our model: the (violable) law that phonological and syntactic boundaries should coincide, and the law that phonological weight units should be as equally distributed as possible, certainly belong to the realm of grammar- so Fodor's (1998) proposal is highly compatible with our view.

Let us conclude this section with a brief discussion of the approach proposed by Phillips (1996). His theory attempts to derive late closure effects from the principle of grammar governing the construction of phrase structure – in fact his approach is quite congenial to ours, with the exception of that emphasis on principles of structure building and the fact that Phillips needs to propose a non-standard grammatical model, in constrast to us.

If a principle such as Phillips' BRANCHRIGHT "Structures should be as right branching as possible" is indeed part of grammar, then it is obvious that late closure effects will be implied by grammar – in fact, this formulation says essentially what was originally designed by Kimball (1973:24) as a parsing principle: "terminal symbols *optimally* associate to the lowest nonterminal node". Unfortunately, Phillips leaves it quite open what the *proviso* in "as right branching as possible" exactly refers to, but the violability of BRANCHRIGHT brings his approach closer to ours in yet a further respect.

We do not want to go into issues such as whether e.g. Active Filler Effects of the data discussed in the next section can be derived from his approach. Rather, we would like to point out that BRANCHRIGHT may be related to a grammatical principle, but it is far from being identical to it. In fact, if BRANCHRIGHT is correct, then something like (61) holds as a default:

(61)   _ c-commands _ if and only if _ precedes _

But (61) is not what is called for in grammatical terms. Kayne (1994) and Chomsky (1995) work with the *Linear Correspondence Axiom* LCA which states, roughly, that an element _ asymmetrically c-commanding _ needs to precede _. C-command relations translate into precedence relations, but the reverse does not hold. Complemeting the LCA in a way such that (61) is stipulated is by no means necessary on any grammatical grounds, and we doubt that the crucial *as .. as possible* can be spellt out in the necessary way. Finally, the LCA was not conceived of as a violable principle, and there is no evidence that it should be conceived as one.

THE EMERGENCE OF THE UNMARKED

That principles of grammar shape the parsing process can be demonstrated in various ways - that these are principles of an OT-grammar makes two particular predictions. What may come to mind first is the import that the ranking of the principles might have on the parsing process - the higher principles should have more influence on parsing than the lower ones. Note it may be difficult to tll such effects apart from a trivial influence of grammatical differences on parsing which each parsing approach must be able to represent, furthermore, as we shall argue below, the relevant situation does not arise easily.

A second prediction from OT parsing is that principles with a (relatively) low rank may nevertheless exert effects on parsing under the appropriate conditions. We concentrate on this second aspect here - it relates to in fact the clearest kind of evidence differentiating OT parsing from other models.

Suppose principle P_x is the lowest of a set of principles $P_1..P_n$ that determine grammaticality in a certain domain D. P_x will therefore have small effects on grammaticality only. But suppose furthermore that the evidence that rules out the effects of P_x comes very late in certain structural types. Then P_x will have a chance of exerting considerable effects on parsing preferences. Nothing comparable can arise in conflict-free grammatical systems. In these, P_x could not be part of grammar. Rather, grammar would contain a specific and surface-true statement P* that is confined to a small array of data. Outside that domain, one would not expect P* to have any influence on grammaticality or parsability.

Consider the assignment/government of Case as a concrete example of this kind. There are, essentially, two ways in which the Case of a noun phrase may be determined, by 'government' or by agreement. Verbs, prepositions or certain functional heads like Infl or Tense may "assign" or "govern" the case of a noun phrase; in more recent grammatical models, one would also say that these heads check the case of a noun phrase. Thus, the verb *unterstützen* "support" assigns/governs/checks the accusative case of its object in German, while *helfen* "help" governs/checks/assigns the dative case.

Furthermore, if two noun phrases stand in a predicational relation, when they are coindexed, Case may be transmitted from one to the other. This is exemplified in (62) for German:

(62) a. er   wird   ein guter   Mann
        he_{nom} becomes a_{nom} $good_{nom}$ man
     b. wir lassen ihn   einen guten Mann werden
        we   let him_{acc} a_{acc} $good_{acc}$ man   become
     c. wir nennen ihn   einen Idioten
        we   call him_{acc} an_{acc} $idiot_{acc}$
     d. er   wird   ein Idiot   genannt
        he_{nom} is an_{nom} $idiot_{nom}$ called

(62a-d) show that the case of a predicate nominal N is a function of the case borne by the noun phrase N is predicated over. Thus, if N is a predicate of the subject (as in 62-a,b), its case varies with the subject case (nominative in a finite clause, and accusative in a causative infinitive). Similarly, N agrees in case with the underlying object in contexts such as (62c,d). In a standard conflict-free grammar, a principle like (63) would be invoked, for which one does not expect any effects on the case assignment to non-predicative noun phrase.

(63)   If NP _ is not referential and predicated over NP _, then _ and _ agree in Case

In an OT grammar, (63) is not, however, the optimal way of capturing what is going on. In some languages Case agreement is a more widespread phenomenon, as (64) from Ancient Greek illustrates: the relative pronoun takes over the genetive case of the head noun of the relative construction instead of realizing the accusative Case required by "possess"- but the reverse would be fine in Ancient Greek as well (see e.g. Harbert 1983). Nothing comparable is possible in

German. The relative pronoun cannot take over the genetive case of the head as in (65), it would have to realize the accusative case governed for it by *sehen* 'see'.

(64) áksioi tes eleutherías hes kéktesthe
      worthy the freedom-gen which-gen you-possess

(65) *wegen des Mannes dessen du siehst
      because the-gen man-gen who-gen you see

From the perspective of Universal Grammar, one will assume two general mechanisms for Case determination, government and agreement:

(66) a. GOVCASE
          An NP must realize the case it is governed for by V, P, etc.
          (Or, rather: there must be a Spec-head agreement relation for Case between NP and
      a head)
      b. AGRCASE
          If NP1 and NP2 are (semantically) coindexed, they agree in Case

In German, GOVCASE definitely dominates AGRCASE, so that the latter principle has a chance of manifesting itself only if there is no Case governor, as seems to be true for the predicative nominals in (62). In Ancient Greek, the two principles seem to be 'tied', i.e. neither dominates the other, so that governed case may be retained or give way to an agreement case. These considerations lead us to an empirical prediction concerning the configuration (67)

(67) .... [$_{NP}$ N, case= _ [$_{RelC}$ rel pronoun case=_ [.......... V]]]

Suppose a relative clause construction such as (67) is encountered, and suppose the case of the relative pronoun is locally ambigous. AGRCASE will establish a preference for _ =_, a preference that will be overridden when verbal case government requires _ to be different from _ , because GOVCASE dominates AGRCASE. But notice that such a conflict can be detected very late only in German relative clauses, because these are verb final, so that information concerning the choice of the governed case comes in very late. Thus, we predict there to be an effect of AGRCASE in the processing of relative clauses, although the grammar tolerates no overt manifestation of case agreement in this area. As we have argued above, similar expectations do not arise in other models of grammar.

To be more precise, consider the initial three elements of the NP+relative clause construction in (68). Feminine articles, nouns and relative pronouns are case ambiguous (if they are subjects or direct objects), but suppose that other grammatical information has disambiguated *die Frau* for e.g. the accusative interpretation. AGRCASE then implies that this accusative case should be transmitted to the relative pronoun. Thus, in contrast to what holds for other construction types involving locally ambiguous initial elements (questions, declaratives), we should be able to see an object preference for the relative pronoun *die* under certain conditions.

(68) die Frau die
      the woman who

This line of reasoning presupposes, of course, that the factor(s) triggering a subject preference otherwise are outranked by AGRCASE. It is not too easy to show this, because AGRCASE hardly interacts with the other principles in question. At least, we may observe in general that respect for Case rules is highly esteemed in German grammar in general – they override the EPP (in subjectless constructions like *mich friert* "me freezes, I am cold) and the parsing evidence suggests that the "Case block" overrides OBLHD, too (the need to introduce a Case assigner/checker always overrides OBLHD-considerations). We know of no other observations that would imply that AGRCASE must instead be ranked below the Extended Projection principle, so we assume that it is not. We know of know consideration that would force the ranking of the "Case block" below any other relevant principle.

Consequently, when (68) is parsed and *die Frau* has already been identified as an accusative noun phrase, (66b) implies that the relative pronoun *die* bears accusative Case as well. But respecting (66b) does not exempt a phrase from the need to respect (66a), as well, that is, the parser will construct a structure down to VP, in order to satisfy the BIJECTION PRINCIPLE as well as (66a). In other words, an object preference is predicted for the relative pronoun in case the target noun bear asccusative case.

The following three experiments test this hypothesis. Experiment 1 is taken over from Schlesewsky (1996), the other two are follow-up studies not reported elsewhere.

Experiment 1: Case agreement effects in German relative clauses
Method
• Participants
20 students of the university of Potsdam participated. They were native speakers of German, and not not familiar with the purpose of the study. They were paid for participation, or received credits.

• Procedure
Subjects read the experimental material in a self-paced reading study with non-stationary presentation and phrase by phrase retrieval. The segments for phrasewise retrieval are indicated in Table 1. Sentences ended with a punctuation mark. After the presentation of the punctuation mark, the participants had to carry out a sentence matching task. By pressing a "yes"- or a "no"-button, subjects had to decide whether a control sentence was a *verbatim* repetition of the preceding sentence. The control manipulation did not involve the proper analysis of the grammatical function to the critical phrases: a negation or an adverb could be missing or be added, or a noun could have been changed. This method had proved effective in experiments concerning the processing of locally ambiguous questions (see Schlesewsky et al., in press).

• Material
The experimental items were sentences of the type represented abstractly in (69), and illustrated in (70) below.

(69) NP_1 v_1 [$_{NP2}$ Det_2 N_2 relative pronoun$_2$ adverb NP_3 v_2 aux$_2$] adjunct clause

NP_1 is the matrix subject, followed by the main verb, while the initial three elements of NP_2, a noun phrase modified by a relative clause, were morphologically case ambiguous, because NP_2 is feminine singular. In one version of the matrix clause, the predicate v_1 was *sein* "be", so that Det-2 and N-2 bear (morphologically unmarked) nominative case. In the other version, NP_1 was explicitly marked for nominative case, and the verb was a standard transitive predicate, so that Det-2 and N-2 bear (morphologically unmarked) accusative case. The principled case ambiguity for NP_2 is thus disambiguated by NP_1/ verb$_1$.

The relative clause begins with a case ambiguous relative pronoun followed by an adverb. NP_3 is also case ambiguous, but it differs from relative pronoun$_2$ in terms of number, so that the proper assignment of grammatical functions depends on which NP aux$_2$ agrees with. The relative clause was followed by an adverbial clause, in order to be able to record possible spillover effects. The resulting four conditions are exemplified in (70)

(70) a. Das ist die Frau,   die glücklicherweise   die Soldaten besucht hat, obwohl ..
         that is the woman who fortunately   the soldiers visited has although
     b. Das ist die Frau,   die glücklicherweise die Soldaten besucht haben,obwohl ...
         that is the woman who fortunately   the soldiers visited have although
     c. der Soldat überrascht die Frau,   die glücklicherweise die Männer besucht hat, ...
         the soldier surprises the womanwho fortunately   the men visited   has
     d. der Soldat überrascht die Frau,   die glücklicherweise die Männer besucht haben,
         the soldier surprises the womanwho fortunately   the men   visited have

(70a,c) are subject initial relative clauses; if there is case transmission due to AGRCASE, we expect reading times for the auxiliary to be fastest in these two conditions. There is case agreement between the head noun and the relative pronoun in (70a,d), so reading times on the auxiliary disambiguating the structure should be fastest for these structures if AGRCASE plays a role in processing.

The participants read five experimental items per condition, and were never confronted with two members belonging to a single pair. There were 130 distractor items not involving material analyzable as crucial for the contrast between the conditions. The segmentation for self-paced reading is given in table 1.

Table 2: Segmentation for self paced reading

Matrix subject	matrix verb	Det & N of NP$_2$	Relative pronoun	adverb	NP$_3$	verb	auxiliary	Comp	rest of ad-verbial clause
1	2	3	4	5	6	7	8	9	10

Results

A repeated measures ANOVA revealed there were no significant reading time differences between conditions for segments 4-6. For the following analysis, reading times for these segments could thus be taken as a single variable. The segments following this block were then compared with the means of this block, and reading times for the head noun+determiner. This is illustrated in table 2, which summarizes reading times is ms and accuracy in %. Relevant congruent condition reading times are set in boldface.

	The woman	who fortunately the soldiers	met	have	in spite of	rest of clause	accuracy
	I	II	III	IV	V	VI	
(a) N=nom Rel=nom	806	680	793	**611**	**667**	1146	79
(b) N=nom Rel=acc	773	685	724	736	790	1304	74
(c) N=acc Rel=nom	886	685	720	639	747	1227	74
(d) N=acc Rel=acc	806	677	637	**594**	**608**	1198	76

Segments III - VI were tested against the mean of segments I and II. For the auxiliary, contrasts in the interactions position by matrix clause type ($F(1,19)=6.68$. MSe=338677, $p<.05$), position by relative pronoun case ($F(1,19)=7.40$. MSe=204823, $p<.05$) and position by case agreement ($F(1,19)=9.65$. MSe=164307, $p<.01$) were significant. On the following segment, the first element of the adjunct clause, the interaction position by case agreement was marginally significant ($F1(1,19)=3.57$, MSe=1207870, $p<.08$), with reading times being higher in the incongruent conditions on both segments.

Discussion

Experiment 1 confirms the expectation that can be derived from AGRCASE: if a reading time increase on the auxiliary and the following segment indicates costs of reanalysis, because the initial assignment of grammatical functions to the relative pronoun turns out to be incorrect, then a comparison of reading times for the auxiliary (611 vs. 736, and 594 vs. 639 ms) shows that grammatical functions were prefentially assigned to the relative pronoun on the basis of the case borne by the head noun. (Note that absolute reading times between conditions (70a,b) and (c,d)

cannot be compared because reading times differed as a function of the choice for the matrix construction as such, see Schlesewsky 1996 for a detailled discussion). This effect is expected if AGRCASE exerts its force on language processing, as is predicted by our processing model.

Experiments 2 and 3 were designed to show that the case congruency effect observed in Experiment 1 is really due to case agreement between coindexed categories, and not to other factors. Consider (71) and (72):

(71) a. er hat Schauspielerinnen obwohl er ihnen   traut widersprochen
        he has actresses   though he them-dat trusts contradicted
     b. er hat Schauspielerinnen obwohl er sie   liebt widersprochen
        he has actresses   though he them-acc loves contradicted

(72) a. er hat Schauspielerinnen obwohl er   mir   traut widersprochen
         he has actresses   though he me-dat trusts contradicted
     b. er hat Schauspielerinnen obwohl er   mich   liebt widersprochen
     he has actresses   though he me-acc loves contradicted

When a structure involves a case ambiguous object such as *Schauspielerinnen* in (71), which allows both a dative and an accusative reading, serious garden path effects arise if this noun is coreferent with a pronoun in an adverbial clause that intervenes between the noun and the pertinent matrix verb, if the pronoun bears explicit case morphology, and if the case expressed by the pronoun is not the one that the matrix verb governs. Thus, (71b) seems more difficult than (71a), because *widersprechen* requires a dative object, while *sie* bears accusative case.

The effect looks as if it is triggered by AGRCASE. If the case of *Schauspielerinnen* is indeterminate in the first place, and if *Schauspielerinnen* and the pronoun are preferentially interpreted as being coindexed, AGRCASE implies that the case visible on the pronoun should be copied onto *Schauspielerinnen*, in which case the parser runs into a problem in (71b).

Alternative explanations come to mind, however. A reading time difference between (71a) and (71b) might also be due to a priming of case frames. It is therefore crucial to check whether such a difference in reading times has a chance of showing up in (72) as well, a pair of sentences that shares the case frame difference with (71), but lacks a coreferent pronoun. Thus, by a comparison of (71) and (72), the role of coindexation for Case transfer can be identified - which would not be possible for relative clause where the two categories involved corefer obligatorily.

Experiment 2:
Method
• Participants
Twenty-four undergraduate students were paid to participate in the experiment. All subjects were native speakers of German and were not familiar with the theoretical background of the experiment.

• Procedure
As in Experiment 1

• Material
The experimental items were sentences of the type represented abstractly in (73), compare also (74). The sentences begin with a matrix clause, which is followed by a complement clause, into which an adverbial clause is embedded.

The crucial NP_2, the object of verb-2 (of the higher complement clause), is a determinerless feminine plural noun phrase case ambiguous between a dative and an accusative interpretation. The pertinent verb $verb_2$ either accepts an accusative or a dative object. Between NP_2 and verb-2, an adjunct clause intervenes that contains a feminine plural object pronoun that can appear with either dative or accusative case morphology, depending on what the adjunct clause verb, $verb_1$ requires. Case morphology of the pronoun is explicit. The pronoun agrees with NP_2 in person, number and

gender, so that the two can be understood as being coreferent, i.e., coindexed in more technical terms.

(73)   matrix clause material +
        COMP NP$_1$ NP$_2$ [conjunction NP$_3$ pronoun$_4$ verb-1] verb-2

The experimental items could thus appear in four conditions: matrix accusative verb (74b,d)/ matrix dative verb (74a,c) vs. complement accusative pronoun (74a,d) /dative pronoun (74b,c).

(74) Der Anwalt bezweifelt, dass man Polizisten,
     the attorney doubts   that one policemen
       a. obwohl man sie   einlädt, glauben   sollte
          though one them$_{acc}$ invites believe   should
          "the attorney doubts that one should believe policemen, though one invites them"
       b. obwohl man ihnen   vertraut, bestechen   sollte
          though one   them$_{dat}$ trusts   bribe   should
          "the attorney doubts that one should bribe policemen though one trusts them"
       c. obwohl man ihnen   vertraut,   glauben   sollte
          though one them$_{dat}$ trusts   believe   should
          "the attorney doubts one should believe policemen that though one trusts them"
       d. obwohl man sie   einlädt,   bestechen sollte
          though one they$_{acc}$ invites   bribe   should
          "the attorney doubts that one should bribe policemen though one invites them"

The participants read **five** experimental items per condition, and were never confronted with two members belonging to a single pair. There were **134** distractor items not involving material analyzable as crucial for the contrast between the conditions.

Results
Data from all 24 subjects were used to the analysis. Only correct responses were included in the analysis of variance. Table 3 shows reading times for the final segments and the accuracy for the sentence matching task. A repeated measures ANOVA showed a significant reading time difference on the mean of both clause final segments; $F1(1,23)= 20,78$, $p< .01$; $F2(1,19)= 16,74$, $p< .01$. The verbal complex is read faster in sentences where the Case of the object noun and the pronoun agree (74c/d).

condition	NP 3	Pronoun 4	Verb adjunct clause	verbal cluster	accuracy
matrix: dat compl: acc (74a)	812	675	611	632	83.4
Matrix: acc Compl: dat (74b)	809	675	618	574	82.7
Matrix: dat Compl: dat (77c)	714	670	628	542	82.9
Matrix: acc Compl: acc (74d)	747	652	649	529	83.5

Discussion
The results of Experiment 2 show the expected effect of case agreement on reading times. Reading times for the verbal cluster are higher in the incongruent condition than in the congruent one. This suggests that the case of the pronoun has been transferred to the coindexable noun phrase, which implies a later need for reanalysis in the incongruent condition.

Case transmission is not the only way of explaining Experiment 2, though. Alternatively, the case frame of the adjunct clause verb may have triggered the expectation of an identical verb type for the matrix clause. Reading times might then have gone up when this type of expectation was not fullfilled. In order to control for this possibility, we conducted Experiment 3 that differs from Experiment 2 only insofar as the explicitly marked pronoun could not be coindexed with the locally ambiguous noun of the matrix clause, because it was either first or second person. This change should not affect case frame expectations based of verb classes, but if the case transmissions observed so far depend on case agreement between coindexed phrases as implied by AGRCASE, we should now fail to see any congruency effects.

Experiment 3:
Method
• Participants
Twenty high-school students were paid to participate in the experiment. All subjects were native speakers of German and not informed about the purpose of the experiment.

• Procedure
As in Experiment 1

• Material
The experimental material used in Experiment 3 differed from the one used in the preceding experiment only insofar as pronoun 4 could no longer be coindexed with NP2, because it was a first or second person pronoun. The participants read six experimental items per condition, and were never confronted with two members belonging to a single pair. There were 134 distractor items not involving material analyzable as crucial for the contrast between the conditions.

Results
Data from all 20 subjects were used to the analysis. Only correct responses were included in the analysis of variance. Table 4 shows the reading times on the final segments and the accuracy for the sentence matching task. A repeated measures ANOVA showed a significant main effect for Case on the final verbal cluster; $F1(1,19) = 9,42$, $p < .01$; $F2(1,22) = 6,32$, $p = .02$. As shown in table 4, reading times are faster when the final verb licenses an accusative object. Reading times for the final verbal cluster are slower in sentences with a dative pronoun in the intervening adjunct clause. Finally, the accuracy was better for sentences where the object case in the embedded clause and the case of the pronoun were identical; $F1(1,19) = 10,15$, $p < .01$.

Table 4

condition	NP 3	pronoun 4	verb adjunct clause	verbal cluster	accuracy
matrix: dat compl: acc	650	534	561	699	91
matrix: acc compl: dat	671	528	567	696	89
matrix: dat compl: dat	653	537	563	799	97
matrix: acc compl: acc	670	531	571	644	95

Discussion
In contrast to Experiment 2, there was no significant reading time difference concerning the case agreement parameter for the verbal cluster in Experiment 3. Given that Experiments 2 and Experiment 3 differed in terms of the coindexability of the pronoun used in the adjunct clause only, we conclude that the factor that explains whether case transmission between two positions applies or not is whether they are interpreted as being semantically related. If they are, AGRCASE implies agreement for case between the two nominals as a default.

PARSING DIFFERENCES

The preceding section has shown that a Case principle normally outranked by other Case constraints may have considerable effects on parsing, because the information relevant for the conflicting principle with a higher rank comes in late only. The prediction that rules which are restricted to small domain in grammatical terms may have considerable parsing effects seems to be made by an OT parsing approach only, so that the data discussed above seem to constitute considerable evidence in favor of our approach.

Quite in general, parsing preferences arise when violable principles that are applied "too early" in parsing and have a chance to favor one analysis over the other. Parsing preferences thus tend to rely on the *absence* of applying the ranking. We thus expect a certain interlinguistic stability of parsing preferences, since grammatical <u>differences</u> are coded in the ranking, which not crucially applicable at the beginning of the clause. In this respect, our approach does not appear to be too far off the track, but one should interpret the availbable evidence with care: our evidence for an initial subject preference comes from subject-initial languages with a nominative-accusative case marking.

There is a certain conspiration of grammatical principles in German and similar languages favoring a subject analysis for a clause initial noun phrase. This need not hold for other languages. Thus, the subject case nominative of German is alsos the one that imposes the leasr amount of conditions on sentence structure: it is assigned as the default in the presence of a simgle Infl node. In ergative/active languages, however, absolutive Case is the default, while subjects of transitive constructions (and certain intransitive ones that may be underlying transitives, though) bear the marked case ergative that is assigned only in a full clausal structure with an object, in other words, it may be assigned only if one assumes two case checking heads. Thus, when a clause initial noun phrase is case ambiguous (e.g. because it is long and the case marker is the last element), the EPP may still favor a subject (=ergative) interpretation, whereas OBLHD now seems to favor the absolutive = (deep structure) object interpretation. For obvious reasons, the ranking of the two principles now IS crucial. Several such constellations are conceivable (e.g., in the VS/VOS languages, the EPP implies a subject preference for an NP following even when the verb is transitive, while, say order rules concerning the placement of specifiers favor the object reading, etc.), but we seem to lack the crucial experimental evidence.

One domain in which variation in parsing preferences and grammar appears to go hand in hand is the fate of the subject preference in the case of an extraction out of a finite complementizer-introduced *that* clause. According to Frazier (1993), there is a preference for the object interpretation of a wh-phrase for Dutch clauses involving long movement, in contrast to simple clauses. Reading times are longer (1209ms) when the verb morphology of the complement clause disambiguates the locally ambiguous wh-phrase for the object-interpretation (as in (75a)) than in the case of a disambiguation towards the object reading (1113ms, as in (75b), cf. Frazier 1993)).

(75) a. welke patient meende Jan dat t de doktors bezocht
       which patient believed Jan that the doctors visited$_{sg}$
       "which patient did John believe visited the doctors?"
     b. welke patient meende Jan dat de doktors t bezochten
       which patient believed John thatthe doctors visited$_{pl}$
       "which patient did John believe that the doctors visited"

Frazier (1993) relates this shift in preferences to the "classical" Empty Category Principle (ECP) (Chomsky 1981): traces of movement have to be "properly governed", that is, they must be governed by a verb or an intermediate trace in the specifier position of CP (to simplify matters a bit). If the parser ignores these intermediate traces in the first pass (as Frazier assumes), the preference shift follows from the grammar, since the subject trace in (75a) violates a grammatical principle then - the ECP. In this respect, Frazier's reasoning comes close to the central idea we pursue here.

Note first, however, that parsing preferences for short and long movement do not always differ. In particular, in studies that have been carried out in collaboration with us (see Schlesewsky et al.

1996), speakers of (Southern) German showed a subject preference for long movement construction in structures such as (76) as well:

(76) a. welche Frau   denkst   Du   dass   der   Mann   gesehen   hat
          which woman   think   you   that the_{nom} man   seen has

       b. welche Frau   denkst   Du   dass   den   Mann   gesehen   hat
          which woman   think   you   that the_{acc} man   seen has

Dutch and (Southern) German thus differ substantially with respect to parsing preferences for long movement. How can this be accounted for? Although the grammars of the two languages are closely related, Dutch differs from German in terms of the identification of grammatical functions, which implies a number of further subtle or coarse syntactic differences, see e.g. Haider (1993). Thus, Dutch has true subject expletives (*er "there"*), whereas German does not. Following Grimshaw's (1997) analysis of *that*-trace effects in English, we can understand the situation in Dutch if we assume that the principle that traces must be lexically governed (TLEXGOV, Grimsawh 1997:416) has a higher rank than the OBLHD. The subject preference for moved elements will thus arise only if lexical government of the trace is guaranteed, which is not the case in clauses with a complementizer, but which arguably holds if the verb has moved up to Comp, as is the case in Dutch and German main clause.

There is little evidence that TLEXGOV and OBLHD should be ranked differently in German. What one observes with (76) rather may reduce to the fact that it is "easier" for subjects to be lexically governed in German than in Dutch. According to Koopman & Sportiche (1991) and much subsequent work, all subjects orginate in the verb phrase, either in the specifier position of transitive (77a) and intransitive (77b) predicates, or as complements for unaccusative verbs. The subject then moves to the specifier position _ of a TensePhrase, and InflPhrase, the Sentence (whatever term you prefer), for case reasons, or because of respect for the EPP.

(77)   a. $[_{TP} \_ \text{ Tense } [_{VP} \_ \text{ } [V \_]]]$
        b. $[_{TP} \_ \text{ Tense}[_{VP} \_ \text{ } [V]]]$
        c. $[_{TP} \_ \text{ Tense}[_{VP} [V \_]]]$

This movement to the specifier of TP is, apparently, obligatory in English (the EPP has a high rank), less obligatory but still preferred in Dutch, and possible but by no means preferred in German. If _ is lexically governed in all three subcases of (77a), as we assume, the difference between German and Dutch *parsing* reduces to the well-motivated assumption that the default for a subject position is in VP in German but in Spec,TP in Dutch, so that TLEXGOV has a chance to override preferences in Dutch, but not German. The question then of course arises whether we do not want to identify the projections of Infl and the Verb in German, as proposed by Haider (1993). We will leave this issue open here.

LIMITATIONS OF THE APPROACH
Everyone would agree that grammatical principles should have a chance of contributing to processing difficulty and parsing preferences. Our claim is that they are sufficient to account for the *formally* triggered parsing preferences. We have identified two problematic cases above (relative clause preferences arising in because clauses, and perhaps some late closure effects, but the phonological link proposed by Fodor 1998 suggests we are on the right track here, too), and quite a number of cases in point.

We do not claim, however, that other factors cannot affect parsing, too (an idea that would be false on obvious grounds). As long as syntax proper is only indirectly involved, no problem arises.

Relative clause attachment preferences seem to be a good example for preferences that are not syntactically triggered. With one exception to be discussed below, it does not seem to be very likely that there are principles of syntax that favor an attachment of the relative clause to either of the two NPs in (78). The clear and lucid discussion in Frazier & Clifton (1996) shows, we believe, that this is a correct prediction, because attachment preferences are influenced by a set of factors alien to syntax.

(78)   the servant of the queen who was shot

As Frazier & Clifton (1996) argue, the issue of which noun phrase is in fact modified by the relative clause need not even be linked to a structural attachment, quite in line with what at least one syntactic approach implies anyhow (Culicover & Rochemont 1990). The failure of our model to predict attachment preference is thus a wellcome result.

Closer consideration suggests, however, that facts may be more complex, however. Taking up an observation that goes back to Sauerland (1996), note that we predict that a relative clause attached in a complex noun phrase should be preferentially associated with NP_k rather than NP_m if the former shares the case of the relative pronoun, but not the latter, see (79) for concrete examples: attachment preferences predicted by AGRCASE are indicated by underlining where they exist. An experiment testing this prediction is under way.

(79)   a.   der   Sohn   des   _Lehrers_ dessen   Tante   …
                the$_{nom}$ son   the$_{gen}$ teacher   whose$_{gen}$ aunt
        b.   der   _Sohn_   des   Lehrers   der   die   Tante ..
                the$_{nom}$ son   the$_{gen}$ teacher   who$_{nom}$ the$_{acc}$ aunt
        c.   der   _Sohn_   von   dem   Lehrer,   der   die   Tante
        the$_{nom}$ son   of   the$_{dat}$ teacher   who$_{nom}$ the$_{acc}$ aunt
        d.   der   Sohn   von   dem   _Lehrer_,   dem   die   Tante
                the$_{nom}$ son   of   the$_{dat}$ teacher   who$_{dat}$ the$_{nom}$ aunt
        e.   der   Sohn   von   dem   Lehrer,   den   die   Tante
                the$_{nom}$ son   of   the$_{dat}$ teacher   who$_{acc}$ the$_{nom}$ aunt

A "perspective sharing principle" in the sense of MacWhinney (1998) "overwrites" case agreement effects in extraposed clauses (see Schlesewsky, Saddy, Fanselow & Kliegl 1998). In an extraposed relative clause, the relative pronoun shows a preference for having the grammatical function of the initial NP of the matrix clause, irrespective of whether the relative clause in fact modifies this initial NP or not. One obvious solution is to assume that there is a locality constraint built into the case agreement process or case agreement checking. If it is true that extraposed relative clause clause are interpreted as if they were independent, the absence of the case agreement effect would be compatible with out model, then. We prefer this view to the idea that a cognitive strategy may override a syntactic principle in a strict sense.

UNGRAMMATICALITY

OT is good at explaining ungrammaticality as triggered by the existence of better competitors, but it is not so good at explaining ungrammaticality in the absence of such competitors ("ineffability"). OT parsing models seem to face a similar problem in this respect: if they are constructed in such a way that they imply the well-known parsing preferences, they must seriously restrict the number of alternatives considered at a given stage in parsing, but then they may fail to note the existence of a 'better' competitor, that is, they may fail to detect ungrammaticality.

Corrigibility may turn out to be the key for an understanding of the perception of ungrammaticality. Consider e.g. what may happen at the initial segment of a sentence, when a noun phrase _ is encountered. Assume that the case of _ is nominative, or that the morphology of
   is compatible with nominative case. Then, by an application of the principles of OT, a single structural representation will be generated (if we make certain background assumptions): _ will be the subject. Suppose now the next item to be parsed is the verb, which is indeed intransitive, but fails to agree with _ , as in (80a).

(80)   a.   *John walk
        b.   John walks

The resulting representation will still be one for an intransitive clause, it differs from the one for (80b) in terms of an AGREE violation only, but given the input string, the AGREE violation could not be avoided.

(81)   AGREE
        The specifier (of IP, the subject) must agree with its head (Infl, as is manifest on the verb)

The perception of ungrammaticality which we certainly have with respect to (80a) can be reconstructed, however, if the faithfulness of the syntactic representation to the input is a matter of a violable principle itself. Suppose that AGREE is ranked above a principle PARSENUM that requires that the number specification of a verbal category in a syntactic representation is identical with what the morphology of the verb indicates. Then the syntactic representation of (81a) is identical to the one of (81b), the difference residing in a PARSENUM violation for (81a) that (81b) manages to avoid.

We can then hypothesize that a string _ is perceived as ungrammatical if it contains a violation of one of the principles of the PARSE family of constraints that require that the syntactic representation buildup is faithful to the lexical material perceived. Consequently, "ungrammaticality" is a notion quite different from "perception of ungrammaticality", as it should be. First, strong garden path effects give rise to the impression that the sentence is ungrammatical although in fact it is not. Second, ungrammaticalities may fail to be observed, and in a very systematic way so. Thus, Schlesewsky, Fanselow & Frisch (1999) observe that the ungrammaticality that is due to the presence of two nominative noun phrases in a single clause is much harder to detect than the ungrammaticality incurred by double accusatives. In a surprisingly high number of cases, experimental subjects fail to realize at all that a string with two nominatives is not well-formed.

An OT parser makes a number of distinctions that seem to go in the correct direction and help understand this difference in the detectability of ungrammaticality. Thus, double-accusative-NPs-clauses violate the EPP, and thus can be repaired by being unfaithful to the accusative marking of one of the two noun phrases only. Double nominative structures respect the EPP, and if nominative case is not represented as a parsable syntactic feature (rather, it corresponds to the absence of case feature in syntactic terms) the ungrammatical structure does **not** involve a violation of a Parse family constraint in the strict sense: rather, it is an abstract property of the verb (governing accusative) that is not respected. Whether all instances of perceived ungrammaticality can and should be treated along these lines is an open question at the present moment.

REFERENCES

Abney, Steven. 1989. A computational model of human parsing. *Journal of Psycholinguistic research* 18: 129-144.

Alexiadou, Artemis, 1994. *Issues in the Syntax of Adverbs.* Doctoral dissertation, University of Potsdam.

Alexiadou, Artemis, 1997. *Adverb Placement. A Case Study in Antisymmetric Syntax*. Amsterdam: Benjamins.

Altmann, Gerry. 1988. Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes* 3:73-98.

Andrews. Avery. 1982. The representation of case in modern Icelandic. In: J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT-Press. 427-503.

Bader, Markus. 1996. *Sprachverstehen: Syntax und Prosodie beim Lesen*. Opladen: Westdeutscher Verlag.

Bader, Markus, 1997. Syntactic and morphological contributions to processing subject-object asymmetries. Submitted.

Bader, Markus & Michael Meng. 1999. Case attraction phenomena in German. Submitted.

Bader, Markus, Michael Meng & Josef Bayer. 1999. Syntaktische Funktionsambiguitäten im Deutschen – Ein Überblick. Submitted.

Baker, Mark, 1996. *The Polysynthesis Parameter.* Oxford: Oxford UP.

Bever, Tom. 1970. The cognitive basis for linguistic structures. In: J. R. Hayes (ed.). *Cognition and the Development of Language*. New York: Wiley.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.:MIT-Press.

Chomsky, Noam. 1973. Conditions on Transformations. In: S.Anderson & P. Kiparsky (eds.). *A Festschrift for Morris Halle*. New York: HRW. 232-286.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht, Foris.

Chomsky. Noam. 1982. *Some concepts and consequencs of the theory of government and binding*. Cambridge, Mass., MIT Press.

Chomsky, Noam. 1986. *Knowledge of Language*. London: Praeger.

Chomsky, Noam. 1995. *A Minimalist Program for Linguistic Theory*. Cambridge, Mass.: MIT-Press.

Chomsky, Noam. 1998. *Minimalist Inquiries*. Ms., MIT.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads*. Oxford: OUP.

Clifton, Charles. & Lyn Frazier, 1989. Comprehending Sentences with Long Distance Dependencies. In: G. Carlson & M. Tanenhaus (eds.), *Linguistic Structure in Language Processing*. Dordrecht: Kluwer, 273-317.

Crocker, Matthew. 1994. On the Nature of the Principle-Based Sentence Processor. In: Ch. Clifton, L. Frazier, & K. Rayner (eds). *Perspectives on Sentence Processing*. Hillsdale, N.J.:LEA. 245-266.

Cuetos, F,. & Mitchell, D.C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30:73-105.

Culicover, Peter & Michel Rochemont. 1990. Extraposition and the complement principle. *Linguistic Inquiry* 21:23-48.

Deprez, Viviane. 1989. *On the typology of syntactic positions and the nature of chains*. Doctoral dissertation, Cambridge, Mass.: MIT-Press.

Fanselow, Gisbert, 1988. German word order and universal grammar. In: U. Reyle & C. Rohrer (eds.). *Natural language parsing and linguistic theories*. Dordrecht: Reidel, 317-355.

Fanselow, Gisbert. 1990. Scrambling as NP-Movement. In: G. Grewendorf & W. Sternefeld (eds.). *Scrambling and Barriers*. Amsterdam: Benjamins.

Fanselow, Gisbert. 1991, *Minimale Syntax*. Groningen, GAGL.

Fanselow, Gisbert, Submitted. Features, theta-roles, and free constituent order. (Re-) submitted.

Fanselow, Gisbert, Reinhold Kliegl, & Matthias Schlesewsky 1999. Processing Difficulty and Principles of Grammar. In: Susan Kemper & Reinhold Kliegl (eds.), *Constraints on Language*. Dordrecht: Kluwer.

Fanselow, Gisbert & Anoop Mahajan. 1996. Partial Movement and Successive Cyclicity. In: U.Lutz & G. Müller (eds). Papers on wh-scope marking. Working paper 97 of SFB 340, Tübingen. 131-161.

Fanselow, Gisbert & Anoop Mahajan. 1999. Towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity. In: U.Lutz, G. Müller & A. v. Stechow (eds). *Wh-scope marking*. Amsterdam: Benjamins.

Farke, Hildegard. 1994. *Grammatik und Sprachverarbeitung*. Opladen: Westdeutscher Verlag

Fodor, Janet D. 1998. Parsing to Learn. *Journal of Psycholinguistic Research* 339-374.

Fodor, Jerold, Tom Bever & Merrill Garrett. 1974. *The psychology of language*. New York: McGraw Hill.

Frazier, Lyn. 1978. On comprehending sentences: Syntactic parsing strategies. Doctoral dissertation, University of Connecticut.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5: 519-559.

Frazier, Lyn. 1993. Processing Dutch sentence structures. *Journal of Psycholinguistic research* 22: 85-108.

Frazier, Lyn. 1995. Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research* 24: 437-488.

Frazier, Lyn. 1998. Getting there (slowly). *Journal of Psycholinguistic Research* 27:123-146.

Frazier, Lyn & Charles Clifton. 1996. *Construal*. Cambridge, Mass.: MIT-Press.

Frazier, Lyn & Giovanni Flores d´Arcais, 1989. Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28: 331-344

Frazier, Lyn & Janet D. Fodor. 1978. The sausage machine: a new two stage parsing model. *Cognition* 6:291-325.

Gibson, Ted & Kevin Broihier. 1998. Optimality and Human Sentence Processing. In: Pilar Barbosa et al. (eds). *Is the Best Good Enough*? Cambridge, Mass.: MIT-Press. 157-196.

Gorrell, Paul. 1995. *Syntax and Parsing*. Cambridge: CUP.

Grewendorf, Günther & Wolfgang Sternefeld. (eds,) 1990. *Scrambling and Barriers*. Amsterdam: Benjamins.

Grimshaw, Jane. 1997. Projections, Heads, and Optimality. *Linguistic Inquiry* 28:373-422.

Haider, Hubert. 1993. *Deutsche Syntax-generativ*. Tübingen: Narr.

Hale, Ken, 1983**.** Warlpiri and the grammar of non-configurational Languages. NLLT 1, 5-47.

Harbert, Wayne, 1983, On the Nature of the Matching Parameter. *The Linguistic Review* 2: 237-284.

Hawkins, John. 1994. A performance theory of word order and constituency. Cambridge: CUP.

Hemforth, B., 1993. *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass. MIT-Press.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2:

Koniecny, Lars, Barbara Hemforth, Christoph Scheepers & Gerd Strube. 1997. The role of lexical heads in parsing. Language and Cognitive Processes 12: 307-348.

Koopman, Hilda. 1996 (V-movement)

Koopman, Hilda & Dominique Sportiche. 1991. The position of subjects. *Lingua* 85: 211-258

Legendre, Geraldine, Paul Smolensky & Colin Wilson. 1998. When Less is More. In: Pilar Barbosa et al. (eds). *Is the Best Good Enough*? Cambridge, Mass.: MIT-Press. 249-290.

Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Narr.

Lewis, Richard. 1993. An Architecturally-based Theory of Human Sentence Comprehesion. Doctoral dissertation. Carnegie Mellon University.

MacWhinney, Brian. 1998. Perspective taking. Ms. 1998.

Mahajan, 1990. *The A-A-bar distinction and movement theory*. Doctoral dissertation, MIT Cambridge, Mass.

Maruyama, Hiroshi. 1990. Structure disambiguation with Constraint Propagation. 28th *Annual Meeting of the ACL* 31-38.

McElree, B. & T. Griffith. 1995. Syntactic and thematic processing in sentence comprehension. *Journal of Experimental Psychology:Learning, Memory and Cognition* 21:134-157.

Meng, M., 1997. *Preferences and reanalysis in processing wh-questions*. PhD thesis, University of Jena.

Menzel, Wolfgang. 1997. Constraint Satisfaction for Robust Parsing of Natural Language. *Theoretical and Experimental Artificial Intelligence*.

Müller, Gereon, 1997. Partial wh movement and Optimality Theory. *The Linguistic Review*. 14: 249-306.

Müller, Gereon. 1999. Einführung in die Optimalitätstheorie. Ms.

Pesetsky, David. 1987. Wh in situ: Movement and unselective binding. In: A. ter Meulen & E. Reuland (eds). *The representation of (in)definites*. Cambridge, Mass.: MIT-Press. 98-129.

Pesetsky, David. 1997. Syntax and Optimality. In: Diana Archangeli & Terence Langendoen (eds.) *Optimality Theory*. London: Blackwell.

Pesetsky, David, 1998. Some Optimality principles of sentence pronunciation. In: Pilar Barbosa et al. (eds). *Is the Best Good Enough*? Cambridge, Mass.: MIT-Press. 337-383.

Phillips, Colin. 1995. Right association in parsing and grammar. *MIT Working Papers in Linguistics* 26: 37-93.

Phillips, Colin. 1996. *Order and Structure.* Doctoral dissertation. MIT.

Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20:365-424.

Prince, A. & P. Smolensky. 1993. *Optimality Theory. Constraint Interaction in Generative Grammar.* Ms., to appear with MIT-Press, Cambridge, Mass.

Pritchett, Bradley. 1992. *Grammatical Competence and Parsing Performance.* Chicago UP..

Pritchett, Bradley & John Whitman. 1995. Syntactic Representation and Interpretive Preference. In: R. Mazuka & N. Nasai (eds.) *Japanese Sentence Processing*. Hillsdale, NJ: LEA: 65-76.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17: 501-557.

Sauerland, Uni. 1996.

Schlesewsky, Matthias. 1996. *Kasusphänomene in der Sprachverarbeitung*. Doctoral dissertation, University of Potsdam.

Schlesewsky, Matthias, Gisbert Fanselow & Stefan Frisch (1999).

Schlesewsky, Matthias, Gisbert Fanselow, Reinhold Kliegl & Josef Krems (in press). The Subject Preference in the Processing of Locally Ambiguous Wh-Questions in German. In: Hemforth & Koniecny (eds) ...

Schlesewsky, Matthias, Josef Krems, R. Kliegl, G. Fanselow, U. Brueck, & K. Mackedan. 1996. Syntaktisches Priming bei Subjekts- und Objektsfragesätzen. Paper, presented at the 2nd Meeting of the German Society for Cognitive Science, Hamburg.

Schlesewsky, Matthias, Douglas Saddy, Gisbert Fanselow & Reinhold Kilegl. 1998. Case Matching and Ambiguity resolution. Poster, 1998 AmLAP conference, Freiburg.

Schriefers, H., Friederici, A., & Kühn, K., 1995. The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language 34*: 499-520.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Doctoral dissertation. University of Colorado.

De Vincenzi, Marica. 1991. *Syntactic Parsing Strategies in Italian*. Dordrecht: Kluwer.
.