Empty Categories in Sentence Processing

Linguistik Aktuell/Linguistics Today

Linguistik Aktuell/Linguistics Today (LA) provides a platform for original monograph studies into synchronic and diachronic linguistics. Studies in LA confront empirical and theoretical problems as these are currently discussed in syntax, semantics, morphology, phonology, and systematic pragmatics with the aim to establish robust empirical generalizations within a universalistic perspective.

Series Editor

Werner Abraham

University of California at Berkeley
University of Vienna
Rijksuniversiteit Groningen

Advisory Editorial Board

Volume 43

Empty Categories in Sentence Processing
by Sam Featherston

Empty Categories
in Sentence Processing

Sam Featherston

Eberhard-Karls-Universität Tübingen

For Véronique

Table of contents

Foreword

The work reported in this book came about as a result of the realization that the issue of the role of syntactic gaps in processing was unresolved. It is surprising that this should be the case, since there are few fields of study which seem to allow experimental approaches to produce answers to syntactic questions, and these few are generally investigated with great zeal. Gap processing shows the potential to be such a field, and it too had been very popular in the late eighties and the first half of the nineties. The early studies from both self-paced reading and cross-modal lexical priming (first and foremost Nicol & Swinney 1989) had shown clear effects at gap positions and this had been widely accepted as strong evidence that traces, or something similar, played an active part in human sentence processing. This, if confirmed, would be a fascinating discovery, as it would reveal a correspondence between the functioning of the human parser and a construct of generative grammar far closer than is normally assumed. The excitement was deflated with the publication of Pickering & Barry (1991) and their demonstration that the data could be interpreted otherwise, as activation of a complement of the verb at the verb position. Subsequent empirical work such as Nicol (1993) tended to strengthen the impression that the excitement about trace activation had been misplaced, since other accounts, not making use of traces, were available.

Roughly this was the situation when my attention was called to the phenomenon by Harald Clahsen at Essex. While alternative explanations had been put forward, the extensive work necessary to decide between them had not been undertaken, for two main reasons. First, it was difficult to see how this might be achieved using English materials, since objects are normally adjacent to verbs in English, and this was one of the confounds in the data which needed to be resolved. Second, the excitement that psycholinguistics was finding hard answers to questions of syntactic theory had been given a douse of cold water, and disillusionment set in: few doubted that the theoretically less interesting answer of Pickering & Barry would prove to be correct. Harald Clahsen had noticed that, while English data could not distinguish between the accounts,

German materials could, and suggested I should have a look. Thus began a fascinating journey into the various facets of the phenomenon, which, perhaps not surprisingly, became more complex as time went on. The original intention had been to perform one experiment in order to settle the issue of what the previous cross-modal priming experiments had been measuring. Instead of one experiment, three were necessary in order to produce an answer which could not be accounted for by any of the competing theories which do not assume additional processing at gap positions.

There remained, however, the unresolved question of what the data on the same topic from the other methodologies was showing. Surprising results require unanimity in the data to support them: since the cross-modal priming data showed one distribution of effects and the probe recognition data (e.g. McElree & Bever 1989) showed another, the overall position still had to be regarded as doubtful. This led the other experiments reported here, using probe recognition, sentence matching and event-related potentials, as well as self-paced reading. The result of this is perhaps as full an exploration of the topic as could be imagined.

I am occasionally asked why I chose to discuss two different frameworks' analyses of the experimental materials instead of assuming just one as is more usual. The answer is twofold: first, one of the aims of the study was to test the competing analyses of the two grammars. I regard it as deeply unsatisfactory that two generative grammar models can produce such different analyses of the same data. This strikes me as demonstrating that neither model has a sufficiently close relationship with the linguistic data it claims to represent. In order to rectify this situation, I look to psycholinguistic research and corpus-based work to provide a means of adjudicating between them, and this book is in part an attempt to do just that. The second reason for discussing the predictions of both Principles and Parameters Theory and Head-Driven Phrase Structure Grammar is that each of them has its descriptive strengths, but if we wish to gain the maximum insight into a syntactic structure it is often best to consider what each of them has to say about it.

Various other researchers have contributed suggestions, criticisms and revisions to this work, above all Harald Clahsen, to whom I owe many of the ideas for experimentation here. Particular thanks also to Thomas F. Muente and Matthias Gross of the Hannover Medical School. Thanks also to Janet Fodor, Andrew Radford, Claudia Felser, Meike Hadler, Kerstin Maut and Sonja Eisenbeiß, as well as the Psycholinguistics Research Group at Essex for their comments and contributions, both to the design and construction of this work.

Closer to home, I must say thank-you to my parents for their care and support, but most of all to Véronique, to whom this book is dedicated.

Abbreviations

All abbreviations except those ubiquitous in the linguistic literature (e.g. VP) are additionally signalled within the text where they first occur and sometimes subsequently where appropriate.

AgrIOP	Indirect object agreement projection	NP	Nominal phrase
AgrOP	Object agreement projection	p.c.	Personal communication
CMLP	Cross-modal lexical priming	PP	Prepositional phrase
CP	Complementizer phrase	PPT	Principles and Parameters Theory
DAH	Direct Association Hypothesis	RC	Relative clause
DO	Direct Object	RT	Reaction time
DOP	Depth of Processing Account	S	Saturated clausal projection
EC	Empty category	SPA	Semantic Processing Account
ERP	Event-related brain potentials	TP	Tense phrase
GPSG	Generalised Phrase Structure Grammar	TRA	Trace Reactivation Account
		UG	Universal Grammar
HPSG	Head-driven Phrase Structure Grammar	V2	Verb second
		VP	Verb phrase/unsaturated clausal projection
IO	Indirect Object		
IP	Inflexion phrase		

Introduction

The aim of this study is to examine certain assumptions about the way that syntactic structures containing extracted arguments are handled by the human sentence processing mechanism and its embedded grammar. In particular we focus on empty categories (= ECs), phonetically null place-holders for locally absent constituents. We give examples of sentences containing one type of empty category, trace, in example (1) (see Chapter 2 for more detail). In such sentences, it is argued that the displaced element (*who, peanuts, thus far and no further*), while overtly in sentence-initial position, also has a grammatically fully specified but silent copy in its canonical position following the verb.

(1)   a.   *Who* did the boatman ferry [*trace*] across the river?
       b.   *Peanuts* the eminent chef despises [*trace*]
       c.   *Thus far and no further* will I go [*trace*]

While the overt occurrence of the displaced constituent is at the beginning of the sentence, it is thought that the trace enters into local grammatical relationships, such as subcategorisation by the verb. It is further assumed that there is a syntactic dependency between the displaced element and the trace, ensuring that they have identical features. However, while supporters of some grammatical frameworks make great use of ECs in their analyses of structures where categories are displaced or wholly absent, others view them as a theoretically undermotivated construct and prefer accounts which do not require them. Our research aim is therefore to find empirical evidence which might confirm or falsify the existence of ECs in the human sentence processing mechanism.

To do this we contrast the predictions for processing of extraction structures of two models of grammar: the Chomskian Principles and Parameters Theory (= PPT) (Chomsky 1981, 1986, 1995) and the lexicalist Head-Driven Phrase Structure Grammar (= HPSG) (Pollard & Sag 1987, 1994). These are two of the most current theories of grammar in use in linguistics today, but also represent rather different perspectives on the use of ECs. PPT makes use of ECs not only at the canonical location of displaced or absent constituents, but also

in simple sentences, since it appeals to movement as a major explanatory mechanism in quite wide fields of the grammar. The assumption of ECs offers the advantage that it permits other constraints on structure to hold exceptionlessly: for example, unsaturated projections, where more arguments are subcategorised for than are actually present, can thus be excluded by simple rule, which makes for a simple and elegant theoretical structure. HPSG takes another view, preferring not to make use of phonologically null constituents, but locating the additional structure necessary to dispense with ECs at the subcategorizing head. Essentially this extra structure consists of a distinction between local and non-local subcategorisation, with this latter applying to locally absent arguments. This arrangement allows the framework the considerable theoretical advantage of assuming only those constituents which are overtly represented in the input.

The use of ECs is thus an interesting point of contention between syntactic theories, which means that we can test sets of competing predictions. It is the more significant because of the importance that ECs occupy within PPT, for it is a central design feature of this approach that the descriptive facts should not be stipulated in construction-specific rules but rather derived from the interaction of abstract universal principles, which themselves are constraints from outside the language faculty. These principles must necessarily apply without exception, for they are part of the larger mental and physiological system in which the language faculty is embedded. In such a system, ECs provide a mechanism permitting the characterisation of these constraints as absolute and exceptionless. Within PPT, particularly within its most recent instantiation the Minimalist Program (e.g. Chomsky 1995), ECs are thus not an optional feature but rather an essential load-bearing member in the architecture of the grammar. PPT without ECs is therefore not readily imaginable, and a significant portion of the credibility of PPT hangs on their psychological reality. This contrasts strongly with the position of many syntactians working in HPSG, who tend to view them as theoretically undesirable and, while they cannot yet rule out ECs absolutely, are keen to develop analyses which eliminate the need for them. The weighty implications for grammatical theories of firm answers to questions about the psychological reality of ECs thus make the issue challenging and worthwhile.

In this study we investigate ECs with the methodology of psycholinguistic experimentation. In order to do this we must make an assumption: namely that there is some close correspondence between the functioning of the human sentence processor and the mental grammars suggested in the literature. We

need to assume that the sentence processor is attempting to assign to its input the structures which grammatical theories would specify for them (see Chapter 3 for further detail). Given this parser-grammar transparency (cf. Berwick & Weinberg 1984 for discussion of stronger claims), we can generate hypotheses from grammatical theories about the behaviour of subjects presented with sentences containing displaced constituents, and so test the hypothesis that ECs are used in sentence processing. While transparency is an assumption that we cannot strictly test, it is a necessary presupposition within work in the field, because even the most abstract work in syntactic theory must, at some level, base itself on the output of language processing. In fact we have no other way of knowing anything about the mental representation of human language except via the mediation of the language processor. It follows that every syntactic theory therefore is and can only be a model of the structure of the language as implemented by the processor. Transparency therefore is an assumption originally made by syntacticians when they make observations about the grammar from the language data, since this, even if it is introspective data, is nevertheless only accessible via the language processor. There is thus ample precedent for this assumption.

Let us note here that we restrict our discussion of ECs in this work to experimental evidence and do not discuss the theoretical evidence or data from other areas of study such as phonology (see e.g. Sag & Fodor 1995). It is necessary for us to constrain our discussion to a certain manageable field of theory and phenomena, and since we make no appeal to evidence from outside the immediate bounds of syntactic theory and sentence comprehension, we shall not review the argumentation.

The structure of the book is as follows: in Chapter 2 we outline the syntactic background to our work and the use of ECs in PPT and HPSG, and in Chapter 3 we review previous psycholinguistic work on extraction dependencies. Chapters 4, 5 and 6 report our own experimentation, while Chapter 7 attempts to draw a balance of the evidence. In the remainder of this chapter we sketch in somewhat more detail the issues that we address and the experiments that we have undertaken, their design, aims and methodology.

The major question we address in this study is whether there is reliable empirical evidence from behavioural measures which supports the assumption of ECs. A considerable amount of work has been done on this issue and a large body of data accumulated. However the application of psycholinguistic methodologies to syntax is a fast developing and continually changing field, and as more is learned about the processing of sentences, past interpretations can

rapidly be overtaken by newer insights. Because of this and because of certain confounds in the experimental structures, there is as yet no conclusive answer to the question whether there are or are not empirical reflexes of ECs. There are currently four psycholinguistic accounts purporting to be explanations of the data set bearing upon the reality of ECs in the representation of sentence structures, each having its own empirical base and set of predictions for processing (see Chapter 4 for details). One of these, the Trace Reactivation Account (e.g. Swinney et al. 1989; Nicol & Swinney 1989), attributes a range of results showing effects in structures with gaps to reactivation of antecedents by ECs; the others are all traceless accounts, in that they attribute the same findings to other causes, and make no reference to ECs. The Direct Association Hypothesis (Pickering & Barry 1991) holds that experimental effects at gap positions are due to processing of dependents at their subcategorizer and that this has been confused with gap position effects because subcategorizers and gaps are normally adjacent. The Semantic Processing Account (Fodor 1989, 1993, 1995; Sag & Fodor 1995) suggests that the gap-related effects may be due to semantic not syntactic processing, and that they can therefore offer no support to a syntactic construct such as ECs. The Depth of Processing Account (Fodor 1995, p.c.) essentially attributes greater activation to greater processing complexity. Since an extracted argument must undergo more or deeper processing than an in-situ argument, it attains a higher level of mental activation, which causes the experimental effects otherwise attributed to reactivation by trace.

While the predictions of these four accounts are largely indistinguishable in English, the language on which most of this research has focused, we have succeeded in teasing them apart by using materials in another language, German, which differs from English in some key characteristics. There are three key features of German which permit this. Firstly, German has a head-final verb phrase, which means that verbs follow their direct objects. This makes it easier to distinguish between effects at verbs and effects at direct object gap positions. Next, German is a verb second language: finite verbs often appear in second position in the clause, but this causes no such change in order of arguments. This permits us to test direct object gaps which are remote from the overt position of their subcategorizing verb. Thirdly, German allows short scrambling, word order variations within the clause. This permits us to test extraction dependencies which do not cross the verb, which removes another confounding factor from the data. Our experiments use these features of German to distinguish between the competing psycholinguistic accounts of the antecedent reactivation data.

Experiments 1 to 3 are reported in Chapter 4. They use the cross-modal lexical priming methodology and are an attempt to establish whether or not there is experimental evidence for trace, testing cases for which different predictions are made by the different psycholinguistic accounts. We achieved this by testing at the position from which a direct object is scrambled in a verb second structure. For example, in (2) from Experiment 1 the processing of the foot of the dependency of *ihre Münze* "her coin" and its extraction site can be examined remote from the subcategorizing verb, as this is in second position in the clause.

(2)   Die Frau   gab ihre Münze$_i$ dem   Mädchen t_i
      the woman gave her coin   to.the girl

Experiments 2 and 3 differ only slightly and allow us to control for other factors which have previously obscured the differences between the accounts. The three experiments together provide empirical data which discriminates between the four different accounts.

In Chapter 6 we report Experiments 4 and 5. These investigate another subset of ECs, referred to in PPT as NP-trace and PRO. Previous evidence for these constructs is much weaker and more questionable than that for *wh*-trace. We report two experiments on this issue; again the aim is to differentiate between alternative interpretations of the existing data set. The first uses an end-of-sentence probe recognition technique and compares raising, passive, and equi structures against three control conditions to assess whether these ECs, as covert anaphors, produce the same speeded response times as overt anaphoric elements. This design too allows us to test contrasting predictions of the different accounts of previous findings. The second experiment looks at the issue from a new perspective using measurements of event-related brain potentials. The conditions we contrast are raising, equi, and transitive control structures, and the measurements are taken at the point where, on PPT analyses, one structure type has an NP-trace, the next a PRO, and the last no covert category. The aim is to determine whether the structural difference between raising and equi implicit in PPT but absent from traceless accounts is reflected in subjects' brain potentials.

Chapter 6 is our final experimental chapter. This addresses a rather different aspect of ECs, namely, the processing of their binding. The aim is to test the claim that trace binding, and only trace binding, has a particular characteristic. If it has, then it follows that the concept of trace is supported. We employ the sentence matching task, which has been argued to be sensitive to a particular

feature of the structure of the language processor. Specifically it appears to distinguish between trace binding and other grammatical processes, a distinction made in modular theories of grammar and processing but not in integrative ones. We examine the processing of a particular island constraint, namely the one which excludes multiple *wh*-extraction in multiple *wh*-questions. In (3a), for example, we see that a multiple *wh*-question with one extracted *wh*-element *who* is grammatical; (3b) shows that the second *wh*-element *what* may not undergo *wh*-extraction even within the embedded clause; (3c) demonstrates that this must be attributed to some island constraint, since the local string …*did you say what + clause* is legitimate.

(3) a.   Who did you say bought what?
     b.   *Who did you say what bought?
     c.   Did you say what Jack bought?

The key issue is whether the parser distinguishes between local ungrammaticality (4a) and non-local ungrammaticality such as illegal *wh*-trace binding (4b).

(4) a.   *What does you say Jack bought?
     b.   *What did you say Jack bought a parsnip?

Previous results using sentence matching have shown ungrammaticality effects for local ungrammaticality but none for non-local ungrammaticality. This would suggest that the processor does indeed make such a distinction, which would lend support to modular parsing models which permit such features as movement, but here too there is more than one account of the experimental findings so far, and we address a test case.

Conclusions can be drawn from our work at two levels: first it has significance in the debate between different models of generative grammar, since different frameworks make different assumptions about ECs and their processing. While we highlight the implications just for PPT and for HPSG in the text here, our results are equally relevant to other frameworks and should contribute towards resolving some of the differences. At a higher level of generality our findings bear upon the validity of the linguistic as opposed to the psychological approach to the representation of language (e.g. Connectionism (Rumelhart & McClelland 1986), linguistic relativism (MacWhinney & Bates 1987)). If observationally and explanatory adequate models of syntactic processors can be developed on the basis of generative grammars, then this must be seen as a confirmation of the linguistic approach to the characterisation of language as the product of a language faculty, a separate module in the mind with its own

specific structure and qualities. This is in contrast to approaches which treat linguistic behaviour merely as one aspect of wider human behaviour, and assert that the specific characteristics of language are derivable from wider non-module-specific processes. On this level therefore, our work can provide evidence supporting generative grammars as realistic models of the human language faculty.

Empty categories in PPT and HPSG

2.1 Introduction

This chapter contains the syntactic background to our experimental studies on the processing of ECs. We begin with a brief sketch of how the aims and explanatory priorities of PPT and HPSG cause them to hold differing analyses of the same empirical data and reach very different conclusions about ECs. We also outline the variety and internal constituency of the ECs in each of the syntactic frameworks. The assumptions about ECs differ markedly between the two models and form one of the most hotly debated areas of syntactic controversy. Since contrasting predictions about processing can be drawn from these assumptions about ECs, we can test the comparative empirical adequacy of the competing frameworks by comparing these predictions with the experimental data.

In the second part of this chapter we specify certain syntactic assumptions we shall make in our analysis of the structures in our experimental materials. We do not attempt a comprehensive introduction to the two grammatical frameworks we discuss here (see Borsley 1996 for HPSG, Radford 1997 for PPT, and Borsley 1991 for a comparative approach), but rather focus only on those parts of the theories which concern ECs. Within Principles and Parameters Theory we subsume Government and Binding Theory (e.g. Chomsky 1981, 1986a) and the Minimalist Program (e.g. Chomsky 1995). For HPSG we refer chiefly to Pollard & Sag (1987) and (1994). Our view of the models is necessarily broad because the empirical studies we shall discuss span more than twenty years, during which time certain aspects of the frameworks have developed. Since research in processing tends to utilize the most consensual of syntactic analyses, it is rarely necessary to specify in detail which version of a grammar is being referred to in a particular study. This is a productive imprecision, because the differences between earlier and later versions are rarely crucial to the discussion; but we compare earlier and later versions of our two competing grammatical frameworks where this is relevant.

In the final section we outline which variants of each theory we shall adopt

as they apply to the structure of German clauses. This is particularly necessary in our discussion of PPT, for there are a number of assumptions which are frequently adopted within current syntactic discussion but which we shall not take up here. In doing this we are characterising the consensual model assumed by workers in processing (e.g. for German Bader & Lasser 1994; Crocker 1994) rather than making specific stipulations for our own study; in fact these alternative analyses (with one exception) do not change the predictions in our experiments.

2.2   PPT and ECs

2.2.1   The position of ECs in the theory

The position that ECs play in PPT is an important one. Trace theory was developed in the 70s as a mechanism to circumvent some problems encountered in the transformational syntax of the time (e.g. Fiengo 1977; Chomsky 1973, 1981; see also references in Chomsky 1981: 144, fn. 76). Essentially traces served to act as markers for transformations, so that all transformations were non-destructive. If a moved element leaves a trace behind it, the previous structure is still visible to the computational system; this removes the need for the syntax to specify in what order transformations apply. This is important because it permits PPT to be a model of comprehension as much as of production: non-structure-preserving transformations cannot necessarily be run in reverse.

ECs are an essential component of PPT for other reasons too. PPT has always aspired to be a grammatical theory which is not only descriptively but also explanatory adequate (e.g. Chomsky 1965: 24f); that is, it is able to produce a descriptively adequate grammar of every human language using only mechanisms and constructs which are psychologically plausible and realistically learnable. To achieve this Chomsky concludes (e.g. Chomsky 1965: 25ff, 1981: 6ff, 1986a: 51ff) that humans are born with a Universal Grammar (= UG) which specifies quite narrowly what forms the syntax can take. The learning of syntax must thus be reduced to the smallest burden possible consistent with the attested variations in structure between human languages. It is this problem of acquisition which is the main motor behind the theoretical direction that Chomskian grammatical theory has taken. Now this minimal syntax is achieved by the use of *principles* and *parameters*. Principles are universal design features of human language, while parameters are options; they contain variables which

are fixed on the basis of input. In line with the aim of explanatory adequacy, the rule systems are abstract and over-arching; it is the interaction of the general rules which produces what appear to be construction-specific effects.

An important point for us to note is that these principles apply blindly and without exception. A locally absent argument should therefore cause the derivation to crash since the local grammatical constraints are not satisfied. This result is avoided by the assumption that extracted arguments leave behind traces as placeholders and that infinitival clauses have PRO subjects. These ECs permit the principles to hold universally and exceptionlessly and thus be credible candidates for inclusion within UG. The assumption of ECs is therefore necessary for PPT to allow it to aspire to explanatory adequacy.

We illustrate this with some examples here. The Projection Principle (Chomsky 1981: 29, cf. Chomsky 1995: 189) requires that lexical requirements be met throughout the derivation: traces permit this to hold even when constituents are locally absent. For instance, (5a) is ungrammatical because the verb *beat* requires an object. What is more, this object must be immediately following and adjacent to it, as (5b) and (5c) show. (5d) however is grammatical even though the overt local constituents *Hilda regularly beats at tennis* are identical to those in (5a): the assumption that the direct object has left a trace in the canonical object position removes the inconsistency.

(5)   a. *Hilda regularly beats at tennis
       b. *Hilda regularly beats at tennis Dennis
       c. *Hilda regularly Dennis beats at tennis
       d.   Dennis, who$_i$ Hilda regularly beats t_i at tennis, …

The Theta Criterion (Chomsky 1981: 34, 101, 170; 1986b: 86 ff; 1986a: 13 f; 1995: 30 ff) demands bi-uniqueness in theta role assignment: ECs permit this to be fulfilled.[1] In (5d) we understand the *wh*-element to bear the PATIENT role which *beat* has to assign under sisterhood. Trace theory allows a natural account of this: the trace is in the appropriate position to receive the theta role. Binding theory (Chomsky 1995: 92, 1981: 183) too can be simplified in an account which makes use of ECs: in a structure such as (6a) there is no overt clause-mate

---

1.   Theta Criterion: Each argument A appears in a chain containing a unique visible theta position P, and each theta position P is visible in a chain containing a unique argument A (Chomsky 1986a: 97).

antecedent for the reflexive.[2] The assumption of a PRO subject in infinitival complement clauses explains this in a principled manner, that is, without positing a different binding domain for such cases as (6b).

(6) a. Harriet tried to improve herself/*himself
     b. Harriet$_i$ tried PRO$_i^j$ to improve herself^j/*himself

Similar considerations apply within Case Theory (1981: 49 ff; Chomsky 1995: 111 ff, 308 f). Objective Case is normally only checked by complements of verbs and prepositions: *whom* in (7) is in neither of these positions. Again we can provide a ready account of this if we assume a chain as in (7b) with a trace at its foot. Objective Case can be checked at the trace position and shared with its antecedent.

(7) a. Whom did Alfie invite to the dinner?
     b. Whom$_i$ did Alfie invite t_i to the dinner?

Endocentricity expresses the X′-theory notion that every projection has a head, but this would seem to be violated by a structure like (8a): the IP has no local head because of I to C movement. The assumption of a trace with the equivalent features to *have* in I resolves this problem (8b).

(8) a. Have [$_{IP}$ they [$_I$] no shame?]
     b. Have$_i$ [$_{IP}$ they [$_I$ t_i] no shame?]

The development of the Minimalist Program (Chomsky 1995, esp p. 1–11, p. 167–172, p. 219–225) in PPT has brought two more reasons for it to require ECs as an integral part of the theory. One of these lies in the treatment of word order variation between languages: in the Minimalist Program this is brought about by morphological differences between languages. A *strong* feature is one which must be *checked* overtly; a *weak* one can be checked covertly after Spellout, the point at which phonological features are divided from the others. For example, the contrast between French main verbs, which overtly occupy the I position, and their English equivalents, whose overt position is V, is the

---

2. Binding: α *binds* β if α c-commands β and α, β are coindexed (Chomsky 1995: 93).

Binding Principles:

A. An anaphor must be bound in a local domain.
B. A pronominal must be free in a local domain.
C. An r-expression must be free.

(Chomsky 1995: 96)

difference between a strong and weak V-feature in I.[3] Verb raising in French is to check a strong feature and is thus before Spellout and visible to PF. The equivalent features on the English I are weak and can be checked after Spellout, and are thus invisible to PF. In such a system, where movement operations account not only for extraction dependencies but also word order variation between languages, the use of ECs, in particular traces, is unavoidable.

The other reason for the increased importance of trace in the Minimalist Program is the multiplication of movements for checking purposes. All finite verbs and all case-bearing NPs will move at some stage in a derivation in order to check their features.[4] This necessitates some means for them to engage in their local grammatical relationships at their base-generated positions. This necessitates the assumption that all moved elements are members of chains, essentially conduits for the passing of grammatical features, which enable constituents to engage in local grammatical relationships at more than one place simultaneously. All non-head chain members are traces. Within the Minimalist Program, therefore, almost every argument will exist not only overtly but also as a trace, with a chain linking them.

To summarize: ECs play an important part within the structure of PPT. They permit transformations to be structure-preserving, which frees the theory of the requirement to order them. They also allow principles to hold without the many exceptions which would be necessary if a purely surface analysis of sentence structure were adopted. There are clear advantages in explanatory adequacy if constraints on structures can be argued to be absolute and to apply throughout the derivation. Traces thus permit the development of a theory structure which relies upon the interaction of highly abstract principles to produce the observed effects, with constructions such as passive being mere artifacts. Such a theory based on abstract over-arching principles could not be implemented if these principles could not be argued to be universal and hold without exception. It is plain therefore that ECs are not merely an optional addition to PPT, but an important structural member. Their role in the Minimalist Program is if anything more central still. With the drive to make constraints on representations ever more necessary and thus exceptionless, and

---

3.   We shall use IP here to stand for a functional projection between VP and CP where further specification is unnecessary.

4.   We shall refer to nominal projections as NPs, since the internal constituency of nominal projections is not at issue here, noting that Chomsky does the same (e.g. Chomsky 1995). This also has the advantage of being a theory-neutral term.

with forced movement for the checking of covert morphological features being the central explanatory mechanism, ECs become an essential part of the framework.

2.2.2   The empty categories in PPT

In this work we restrict ourselves to the empty categories pro, PRO, NP-trace and *wh*-trace, which have the status of maximal projections and which are obligatory in the structures that license them. We illustrate these in (9).

(9)   a.   Jess seemed (to us) *NP-trace* to like cream best of all.
      b.   What did James want *wh*-trace for Christmas?
      c.   Josh hoped *PRO* to run 100m in less than 10 seconds.
      d.   *pro* nataka   wali na kuku.   (Swahili)
           EC want.PRES rice and chicken

We shall have nothing to say about other empty elements which are only optionally phonetically null or which represent only heads, since the theory relating to them has been much less well developed.

Empty categories occupy a very central role within PPT precisely because they have no overt form (though see Sections 3.3.2.4 and 4.2). This has implications: their nature is very immediately determined by the grammar and much less by external factors; they can be thought of as the minimal element required for convergence, which makes them an interesting keyhole onto the requirements of the grammar. It might even be argued that it is implausible that their existence and characteristics should be acquired by the child exposed to the language, and that they are likely to reflect deeper principles of UG, universal and biologically determined (e.g. Chomsky 1981:55ff).

Chomsky 1981 contains a principled justification of the characteristics and distribution of ECs. Chomsky 1982 (p. 78f, see also Chomsky 1995:41) adds to this the category *pro* ("small pro") allowing the generalisation that the types of nominals, overt and covert, vary by two parameters [anaphor] and [pronominal].

In fact only seven of the eight cells in Table 1 are filled. The overt nominals are anaphors, pronouns and R-expressions; the covert categories are PRO, NP-trace, *pro* and *wh*-trace. The features [Anaphor] and [Pronominal] are most obviously defined by their binding behaviour: anaphors being bound, and pronouns free, in their binding domain. R-expressions, which are neither anaphoric nor pronominal, are subject to neither of these binding conditions. No overt nominal category can be [+Anaphor] [+Pronominal] as this implies