Title	Formatting Lexical Entries: Interface Optionality and Zero
Author(s)	Joseph Emonds
Citation	Theoretical and applied linguistics at Kobe Shoin, No.5 : 1-22
Issue Date	2002
Resource Type	Bulletin Paper / 紀要論文
Resource Version	
URL	
Right	
Additional Information	

Formatting Lexical Entries: Interface Optionality and Zero[*]

Joseph Emonds

Abstract

Transformationalists (including minimalists) currently assume that (i) particular grammars reduce to the store of closed class lexical items, and (ii) syntactic structures project directly from lexical items. But they seem satisfied with common sense specifications of "possible lexical items." This study moves beyond this vague pre-theoretical stage, focusing on how best to lexically notate *optionality* and *null realizations* at both the PF and LF interfaces. It argues that both the symbol ∅ and the parenthesis notation express linguistically significant generalizations in each of the phonological, syntactic, and contextual parts of lexical entries. As discussed here, their proper definitions allow us to construct simple and in principle easily learned lexical entries which fully explain many alternations between null and non-null PF allomorphs and between distinct yet partly similar interpretations of other grammatical morphemes such as English *of, to, there, from, it, -ing, -en* and Japanese *–(r)are*.

1. Some issues in lexical formalisms

One can imagine a grammatical model for natural language in which a lexicon plays little or no formal role. Chomsky (1957) made a revolution in linguistics by presenting such a system. But current models of natural language, for example that of Minimalism in Chomsky (1995) and succeeding works, now crucially depend on the form and content of lexical entries, basing themselves on concepts such as lexical arrays and the strong features of various grammatical items.

In fact, the dependence of recent transformational models on the lexicon is almost total. Their advocates currently typically assume that (i) particular grammars reduce to nothing more than the store of closed class lexical items in a language, and (ii) all syntactic structures project directly from lexical items. These claims sound very restrictive and scientific—until one realizes that most of these syntacticians are operating with essentially no theory about the form of closed class lexical items.[1] There is not even any agreed on model of mechanisms for the more widely studied open class lexicon; lexical semantics is usually

[1]This state of affairs has a rather dismaying implication: transformational generative grammar studies structures that are fundamentally intuitive and unformalized. In studies spanning more than 40 years, one could hardly find for all languages taken together a dozen closed class items whose lexical entries have been analyzed in terms of

treated as some kind of peripheral concern, orthogonal to serious syntactic theorizing. For the most part transformationalists—in particular practitioners of both government and binding and minimalism—are satisfied with only the broadest and vaguest common sense specifications of "possible lexical items." This study will try to move beyond this state of confident satisfaction with an empty lexical theory.

Of course, we must start with the truism that a lexical entry's feature specifications are indeed phonological π, syntactic σ and semantic λ. Conventionally they are written left to right in this order, e.g., for a manner of speaking verb, something like: *murmur*, V, ACTIVITY, Speech, Not Loud, Indistinct.

But let me begin by pointing out a discrepancy lurking in these simple formulations. Although a priori discussions often treat phonology and semantics as parallel, we must keep in mind that syntactic notation σ is not "neutral" between the categories of π and λ. On the one hand, syntactic features σ are entirely disjoint from the phonological features π used in Phonological Form ("PF"). On the other hand, a language's syntactic features and categories σ are *not* distinct from the features used at Logical Form ("LF"). Rather, σ are typically (and in my view always) drawn from of a fixed subset of $\Lambda = \{ \lambda \}$, i.e., syntax is both disjoint from phonology and also (in my view completely) "cognition-based." For this reason, I call the σ "cognitive syntactic features."

This syntactic subset of Λ, whose members I notate as F, is largely universal—a version is the bar notation with a fixed number of heads—though particular languages may syntacticize somewhat different sets of semantic features λ. The many semantic features which are not part of the syntax of a language are notated here f, i.e., $\{ f \} = \Lambda - \{ F \}$. I argue in Emonds (2000) that no lexical classes except N, V, A and P ever utilize members of $\{ f \}$, and that the syntactic behavior of items with f (open class elements) greatly differs from that of items with only F (closed class elements).

Additionally, at least the lexical head categories N, V, A and P have contextual features—a variant of subcategorization as introduced in Chomsky (1965, Ch. 2) extended to word-internal frames as in Lieber (1980). Emonds (2000) argues extensively that relatively minor revisions of the original subcategorization mechanisms are the best current model for these lexical contextual features. Certain of these revisions concern selection of phrasal complements.

First, I incorporate the widespread realization that individual phrasal subcategorizations should not specify left-right order; hence I replace the more classical notations +__XP and + (XP) with the order free notations ⟨XP⟩ and ⟨(XP)⟩. Lieber's word internal frames then should be written ⟨Y__⟩ for suffixes and ⟨__Y⟩ for prefixes. These new enclosing symbols ⟨ ⟩ *always symbolize contextual* not inherent features.

Second, I argue that lexical items are listed to select (lexicalized) *heads* of complements and features of those heads. A classical feature +__DP thus becomes ⟨D⟩; in general, lexical entries simply never mention phrases. I adhere to this notation here throughout.

Using this result, Emonds (2000) then claims to establish the following restriction on lexical items:

(1) *Semantic Atomism. Lexical specifications of contexts use only cognitive syntactic features F.*

---

well-defended and sophisticated formal proposals. Put simply, transformational grammar has not yet succeeded in convincingly characterizing a significant number of grammatical morphemes in even a single language.

That is, contextual features exclude any purely semantic mechanisms stated in terms of λs, such as theta grids or lexical conceptual structures, which are not independently justified as syntactic features F. This restriction (1) vastly simplifies the lexicon, i.e. the component that a child must learn based on experience.

The claims that syntax is cognition-based and that contextual features are all syntactic have consequences for lexical notation. For perspicuity let us divide the four parts of a lexical entry by slashes as in example (2): the first part is phonological, the second syntactic inherent features, the third syntactic contextual features, and the fourth inherent (non-contextual) purely semantic features.[2] The second and third parts are then entirely in terms of {F}, where the F are all λ drawn from Λ.

(2) murmur / V, ACTIVITY / $\langle\langle$ ([D, DATIVE]) $\rangle\rangle$, $\langle$ (C) $\rangle$ / Speech, Not Loud, Indistinct

It is far from clear that features can represent the fourth part of entries; an adequate notation may well be different in kind from that for the second and third parts. In any case, the f and/or whatever needs to replace them do not enter into syntactic computations and will not be further discussed in this essay.

Consider now the PF information in the first part of an entry and the LF information in the second. The cognitive basis of syntax leads to an asymmetry in the notions, "not realized at PF" and "not realized at LF." If an item must be phonologically null (e.g., English present tense agreement in marked persons and number), we say that its PF form π is $\emptyset$, without any direct consequences elsewhere in syntax or LF.

There can also be PF morphemes not realized at LF. Some English examples analyzed in some detail in Emonds (2000) include the verb *be*, the gerundive nominalizer *–ing*, the infinitival marker *to*, the complementizer *that* and the preposition *of*. We would like to use the same symbol $\emptyset$ for null at LF as for null at PF. But presumably items that do not function at LF still have syntactic categories σ, which are all λ features as seen above. So in this case what needs to happen is that one (or more) members of the syntactic specifications σ must be *cancelled* at LF. To express this (certainly marked) situation, let us say that the symbol $\emptyset$ in the second part of an entry of an X^0 lexically specifies the category X for elimination at LF. In some other cases, we will see that some grammatical morphemes *optionally* lose their interpretation in terms of their syntactic category X; this will be notated with parentheses as (X). Section 4 justifies in detail this notion of LF cancellation of syntactic features.

An algebraic system such as lexical notation can generally be considered closer to optimal as its notational possibilities become more general—provided that the resulting distributional predictions are all exemplified (justified) by widespread phenomena in natural languages. Questions of whether and how the parenthesis notation or the null symbol $\emptyset$ can be usefully extended to the *phonological or category specifications* in a morpheme's entry seem not to have been examined. But it will be a welcome result if parentheses can surround the material in any part of a lexical entry, and if a single symbol $\emptyset$ uniformly indicates in any part something like "cancelled at an interface."

This essay will thus focus on these issues of *optionality* and *null realizations*. It will argue that descriptive and explanatory adequacy require both the parenthesis notation and the symbol $\emptyset$ in the first, second and third parts of lexical entries for expressing linguistically significant

---

[2] Because of the notation $\langle\alpha\rangle$ and the distinction drawn between the feature types F and f, the slashes are only for convenience.

generalizations. They can fully explain alternations between null and non-null grammatical synonyms and between distinct yet similar interpretations of other grammatical morphemes which otherwise remain mysterious.

2.   Parentheses and ∅ in the Contextual Part of Lexical Entries

As in Section 1, the notation $\langle \beta \rangle$ means throughout, "has a sister phrase whose highest lexicalized head is of category β." So the lexical notations V, $\langle$ANIMATE$\rangle$ and V, $\langle$C$\rangle$ mean respectively, a verb that has Animate DP sisters and a verb that has CP sisters.

If a lexical item in natural language appears optionally in certain contexts, we indicate this in its lexical entry by enclosing those contexts in parentheses ( ), a practice that has remained unchanged since Chomsky (1965, Ch. 2). That is, categories in the contextual part of lexical entries can be parenthesized For example, the data in (3) shows that a lexical entry for *convince* should allow among others the three possibilities in (3a). This is expressed by two sets of parentheses in (3d) around the labels of optional categories in context features:

(3)   a. Mary convinced { her friend / *the situation } ({ that / why } it would be too late).

b.*Mary convinced ({ that / why } it was too late).

c.*Mary convinced a friend for everyone to be on time.

d. convince, V , $\langle$ ANIMATE $\rangle$, $\langle$(C, (WH))$\rangle$, f_k[3]

Parentheses around a context feature β for an item α therefore mean, "β is an optional sister of α at the LF interface."

However, we generally do not use the null symbol ∅ as a context feature to mean "no possible sister for α at the LF interface." Rather, following a convention in Chomsky (1965), we simply leave unspecified all those categories that cannot appear as contexts for a head in a lexical category. Thus, the fact that the manner of speaking verb *murmur* in (2), unlike *convince* in (3), can appear as an intransitive (*Outside John was murmuring*) is indicated in its lexical entry by the fact that *all* its potential complements are parenthesized. Verbs that are purely intransitive (*blossom, exist, snooze*) simply lack context features entirely.

There are, however, different appropriate uses for a lexical contextual specification $\langle$∅$\rangle$.

First, unspecified items in the *functional* categories I and D do have unmarked complements: VP and NP respectively. So the context feature $\langle$∅$\rangle$ for items of these categories can mean "lacks the expected XP complement at both LF and PF." We return to examples in Section 4.3.

Second, predicates in lexical categories that obligatorily take complements of a certain category can contrast, in part arbitrarily, as to whether these complements can be null discourse anaphors or not. Some of these contrasts are first pointed out in Grimshaw (1979); their syntactic nature is discussed in Emonds (2000, Ch. 9).

(4) We don't know when the game starts. You should { ask / find out / *tell / *figure out } (soon).

---

[3]I assume that indirect questions are headed by Cs marked with the feature WH. There are also a couple of other possibilities: *Mary convinced her friend to be on time, Mary convinced a friend of her honesty.*

It may be that the parenthesis notation (α) should be considered a special case of the disjunctive braces notation {α / β}, where β is ∅. I do not treat this question here.

Jim is driving into the office today. Sue will { go / come / *head / *walk / *drop / *get } soon.

I have the house keys. When do you think we can { enter / take over / *occupy / *empty out }?

It appears to be a marked property that some verbs obligatorily subcategorized for certain complement types XP in LF (*ask, find out, go, come, enter, take over*) allow XP to be phonetically null and to be interpreted by virtue of intersentential contexts. They thus contrast with optional complements (marked with parentheses) as in (3d), which seem to be overt at LF if and only if they are overt at PF. Thus, the entries for *find out* and *come* should be essentially (5):

(5)   a.   find out, V, $\langle \{D / C, (WH)\}, \emptyset \rangle$

      b.   come, V, $\langle P, \emptyset \rangle$

This notation with $\emptyset$ means that a complement obligatory at LF can be phonetically null (for *find out* the choice is among DP such as *the answer*, a *that*-clause or an indirect question).[4]

We thus see that both parentheses and $\emptyset$ are useful in the third, contextual part of lexical entries. Parentheses have their traditional interpretation as present or absent and here refer simultaneously to PF and LF. The novel contextual interpretation of $\emptyset$ proposed here, for signaling an "LF discourse anaphor that is null in PF," has in fact been long needed in syntactic representations. These uses are summarized more formally in the concluding section.

3.   Parentheses and $\emptyset$ in the Phonological Part of Lexical Entries

An interesting issue in lexical formalization concerns the extent of null morphemes in natural language. Needless to say, transformational generative grammar operates freely with a conception of empty categories (traces) generated by phrasal or head movement. It also treats elliptic constructions, in which null phrases α are base-generated and co-indexed with some co-referring α under certain structural conditions (cf. Lobeck, 1995, for a fairly comprehensive treatment).

But outside of these non-lexical structures of ellipsis and traces, it is not clear under what conditions other empty X^0 are tolerated. The issue of whether and to what extent optimal linguistic descriptions include optional and/or obligatory null morphemes intrigued structuralist linguists from de Saussure to Z. Harris, but seems to have a secondary status in generative syntax. In generative terminology the question might be put, under what conditions can phonetically zero or "empty" morphemes appear as X^0 projections in surface structure or PF?

One rather obvious restriction on empty X^0 is spelled out in Emonds (2000, Ch. 3 and 4). In terms of the distinction in feature types of Section 1, open class items (proper subsets of N, V, A and perhaps P), characterized by purely semantic features f without a role in syntactic derivations, are distinguished from closed class items (of all syntactic classes) all of whose features F are used in derivations.[5] In particular, there are closed classes of N, V, A and P fully

---

[4] I do not determine here whether the category for $\emptyset$ is DP or CP or either for an entry such as (5a).

[5] This distinction between LF features that play a role in syntax (F) and those that do not (f) makes no provision for a separate class of purely diacritic or formal features. While I explain below conditions under which F (such as PAST, PLURAL, WH, etc.) do and do not contribute to LF, I hold there are no F which never contribute to interpretation. See also below the final note of the essay.

characterized by F (lacking f), as well as the more familiar closed classes of Determiners, Modals, Complementizers, Quantifiers, etc.

With this distinction between f and F, (6) at least seems clear: languages do not tolerate null morphemes as phonetic realizations of open class items.

(6) *A lexical phonological representation π can be $\emptyset$ only if an item contains no purely semantic f.*

Thus, all the null morphemes dealt with in this study are fully characterized in syntax by combinations of syntactic features F_i. (7) provides two simple examples of such morphemes. (7a) is the English present tense number agreement verbal suffix in marked persons and number, where marked feature values are notated by μ:

(7)   a. $\emptyset, V, \mu PER, \mu NUM, \langle\, V\_\_\, \rangle$

       b. $\emptyset, V$

(7b) provides for the null anaphoric VP allowed in English, which contrasts with many languages whose anaphoric VPs must contain a phonetic V. Zagona (1982) suggests that English, as opposed to Romance, is parameterized to permit null VPs, but the assumption that particular grammars reduce to the store of closed class lexical items forces us to recast this parameter (as well as many others) as a lexical property.[6]

3.1   Optionally null grammatical elements in root contexts: the UCP

In order to determine how parentheses are used in PF representations, we must digress to consider some general syntactic conditions on empty categories.

What is often referred to as "mainstream" generative syntax seems to have developed an unfortunate propensity for partially developing answers to grammatical problems, and then abandoning empirical study in favor of a hurried move to supposedly higher levels of explanatory success.[7] This style of research, however, does not exclude the possibility that earlier ideas are the best current basis for improved and more general analyses.

A case in point is research on explaining the distribution of categories empty in PF. Chomsky's (1981) Empty Category Principle ("ECP") requires governed empty categories to satisfy rather stringent conditions of being "properly governed". The government and binding framework additionally allows certain ungoverned categories to be empty without satisfying proper government.

The distinction between governed and ungoverned forms the basis for a possible explanatory account of the deletions exemplified in (8)–(13), which Standard English allows only in root clause contexts:

---

[6]In addition, such empty VPs are well-formed of course only if they are correctly licensed and their content is correctly identified; cf. the huge literature on "VP ellipsis." The point here is that (7b) expresses the brute but simple fact (which a child must easily learn) that English allows them while other languages do not.

It is a curiosity, perhaps a revealing one, that the parenthesis notation would seem to suggest collapsing (7a) and (7b). English is unusual both in having null VP anaphors and in having null verbal inflections in marked persons and numbers. Perhaps this is no accident.

[7]To practitioners aware of the difficulties, this is frustrating on two counts; the theoretical paradigm changes look merely like a strategy for camouflaging lack of success, and possible sharpening and refinements of grammatical analyses are shrugged off as without current interest.

(8) [DP [D ∅]] [I ∅] Been working myself silly lately.
     *The fact that [DP ∅] [I ∅] been working myself silly lately has hurt my family life.

(9) [C[I ∅]j] [IP [DP [D ∅]] tj Working yourself too hard these days? }
     *She asked you [C{if / whether / ∅}] [IP [DP [D ∅]] [I ∅] working yourself too hard these days.]

(10) [DP [D ∅]] [I ∅] Get yourselves a haircut, could you?
     *She approves of my opinion that [DP ∅] [I ∅] get yourselves a haircut.

(11) Everybody [I ∅] prepare yourself for a shock, will you?
     *He will issue a warning that everybody [I ∅] prepare yourself for a shock.

(12) What! ({Them / *They}) [I ∅] be on time? You must be kidding.
     *You claim (that) (them) be on time?

(13) [DP[D ∅] [CP That John smokes a lot]]i [C ∅] [IP Mary finds ti noteworthy.]
     *Mary finds [DP[D ∅] [CP that John smokes a lot]] noteworthy.

The first examples in the above pairs show that the highest heads and phrases in root contexts can be empty in ways that typical embedded heads and phrases (in the second examples of the pairs) cannot be, apparently because the acceptable sentences lack plausible governors for the empty categories. That is, heads and possibly specifiers of root clauses are among the "ungoverned" categories. First, the highest I or C heads clearly have no governors. Second, because these heads are empty, they are in turn taken not to govern the DPs in the SPEC positions. This is corroborated in (12), where a null (deleted or absent) root I does not assign nominative case, which would be a property of a governing I. Since the root clause subjects in (8)–(12) are ungoverned, they are exempt from proper government.

In the topicalized structure (13), a clause is in SPEC(CP) and the highest head is C, which again has no governor. As a result, the head D of the phrase in this specifier of C is not governed either, and can be empty. By contrast in an embedded context this D would have to be spelled out, as in *Mary finds *(the fact) that John smokes a lot noteworthy.* In summary, the patterns of root deletions in (8)–(13) suggest that proper government is imposed on empty categories only in the presence of government.

In fact, this is exactly how the ECP functioned for the limited number of English paradigms that it was used to explain. An empty pronominal subject PRO of infinitives was claimed not to be governed and hence was exempt from the ECP. That is, the ECP expresses the idea that proper government is imposed only on empty categories that are *governed*. Exactly as expected then, the null elements in (8)–(13) can be empty precisely because they are in ungoverned heads and specifiers of root projections; this exempts them from government and therefore proper government and so explains why they can be null.

Although the research around the ECP has trailed off somewhat inconclusively, it seems we should retain the following aspects of research on this topic:

(14) a. The highest heads in trees such as the I and C nodes in the grammatical examples above are not themselves governed.

      b. Is govern subjects (and Cs govern topics) only if they are lexicalized in syntax. Hence, the empty Ds and DPs in the grammatical examples in (8)–(10) and (13) are not governed.

c. Ungoverned categories can be empty without violating the ECP.

d. In contrast, embedded categories typically are subject to the ECP.[8]

In terms of lexical entries, it appears that certain contentful grammatical morphemes such as *I, you, have, are, will, could,* etc. can be empty only in root contexts. That is:

(15) *Ungoverned Category Principle ("UCP"). Certain lexical items with interpreted syntactic features F may be phonologically empty (i.e., $\pi = \emptyset$) if ungoverned.*[9]

Let us return to our main concern, lexical notation. We can indicate which items fall under the UCP by using parenthesis notation around the phonological form in these items' lexical entries as in (16).[10] Since a null allomorph is disallowed in any governed context, no special lexical provision is needed to express this limitation.

(16) a. (I), D, I PERSON, SINGULAR, $\langle \emptyset \rangle$

b. (you), D, II PERSON

c. (have), { I / V}, $\langle [_A -en] \rangle$

The UCP (15) applies as well to the ungoverned null subject pronoun of infinitives PRO. Several grammarians have noted its affinity with the animate pronoun *one* in contexts (17a) where a subject is optionally overt and in others (17b) of obligatory arbitrary PRO.

(17) a. It is not unusual (for one) to take { oneself / *himself } too seriously.
This shower is modern enough (for one) to clean { oneself / *itself } without much trouble.

b. How to clean { oneself / *himself / *itself } is unclear.

While previous analyses of base-generated PRO have failed to capture this rather obvious equivalence, our new use of the parenthesis notation easily expresses the alternation of *one* with PRO.

(18) (one), N, ANIMATE, −SPECIFIC

As expected from a combination of (15) and (16a–b), ungoverned PRO subjects can also have first or second person features in English:

(19) It is better to take { myself / yourselves / oneself / *herself / *themselves } seriously.

---

[8]The next subsection returns to a lexicalization condition on the governor required under proper government.

[9]A topicalized first or second person pronoun requires stress and thus cannot be realized as a zero allomorph: *You we think have been working yourself too hard,* but *$\emptyset$ we think have been working yourself too hard.*

The UCP (15) is not the only condition under which interpreted sets of grammatical F_i can be phonologically empty. Sets of features F can be $\emptyset$ in PF if these F are"alternatively realized" on a neighboring head; this is the condition that regulates when subjects exhibit null pronoun subjects in "pro-drop" languages. Null root subjects in English do not result from pro-drop. Cf. Emonds (2000, Ch. 4) for details.

[10](16) oversimplifies somewhat the conditions in English root clauses required for null allomorphs; Hendrick (1982) is a more extensive study.

The UCP applies only to the lexical items that are specified for empty ungoverned allormorphs. That is, the entries of the ungoverned categories in (8)–(11) specify that only these particular morphemes (*I, you, could, will*), as opposed to their near relatives, are possibly empty.

(20) I just talked to my parents. *Been working themselves silly lately. (*they→ $\emptyset$)

Have you seen Sue lately? *Working herself too hard these days? (*she→ $\emptyset$)

*Get yourselves a haircut, might you? (*might→ $\emptyset$)

*Everybody prepare yourself for a shock, may you? (*may→ $\emptyset$)

Moreover, these forms sometimes must be specified as possibly empty only in certain contexts. For example, forms of *will* and *can* cannot ordinarily delete:

(21) You *(could) stay out late last month, couldn't you?

Under socialism, everyone *(would) get a living wage, wouldn't they?

But *can* and *will* may be left unpronounced when accompanied by an imperative feature IMP in main clause (UCP) contexts:

(22) Could everyone kindly get a haircut?

Everyone kindly get yourself a haircut, (could you?)

You stay at home please, (won't you?)

The possibility of conditioned null I can be expressed in lexical entries by first parenthesizing the phonological material and then linking these parentheses via indices to whichever feature conditions the potential silent alternative:

(23) (could)$_i$, I, MODAL, POTENTIAL, { PAST / IMP$_i$ }

This entry is to be read: *could* is a modal expressing potential and optionally either the syntactic PAST or an imperative wish / command with IMP. The co-indexing shows that only in the latter case is its phonological form optional. Such silent but contentful I surface only when allowed by the UCP (15). That is, *could* alternates with $\emptyset$ only in a root IP: *He asks whether everyone*(could) please get a haircut.*

Another example of a structurally conditioned empty allomorph in a root context is the copula *are*. Standard English does not tolerate an unpronounced allomorph of uninverted *are*:

(24) Well, as for us, (we have) just been working ourselves silly lately.

*Well, as for us, (we) just working ourselves silly lately.

(You have) been very impolite, haven't you?

You *(are) being very impolite, aren't you?

It thus appears that the parentheses around *are* must be conditioned by it being in inverted position, i.e. in C, as in the lexical entry (25). That is, *are* is optional if and only if it is in C and is additionally Second Person.[11]

---

[11] The optional presence of *are* in C should perhaps be instead expressed by a parenthesized context feature ⟨(I)⟩. Section 2 defines parentheses around the first or primary syntactic category feature, but not around subsequent ones. Hence we can define parentheses differently with secondary features such as C in (25). I don't explore this further here.

Third person *are* lacks a null allomorph in Standard English: **(How) { all of them / your boxes } being shipped?*

(25) (are)ᵢ, I, (C)ᵢ, −PAST, PLURAL, (II PERSON)ᵢ

The entry (25) is to be read: *are* is an I with the features −PAST and PLURAL; when it additionally has the features (categories) C and II PERSON, it may be left unpronounced, subject of course to the UCP (15). Other silent morphemes in root contexts are doubtlessly subject to similar somewhat complex lexical specifications. The effect of the UCP is to uniformly exclude these null allomorphs in embedded structures.

A final question on the UCP concerns the level(s) in a derivation when the empty allomorphs it sanctions may appear. In other work such as Emonds (2000), I claim that phonological realization actually operates in LF (as well as PF) to signal interpretability, except in cases of co-indexed elements such as traces and ellipsis. Therefore, parenthesized phonology for interpreted elements, as in (16), (23) and (25), implies not a failure to insert phonological material but rather its absence at only the PF interface. In classical transformational terms, these null allomorphs result from late deletions rather than from non-insertion.[12]

The UCP, which grows out of the ECP of government and binding, is thus the basis of my first extension of parenthesis notation: the phonological part of a lexical entry can appear as (π), as in (16), (18), (23) and (25). This allows the phonological form π of interpreted grammatical morphemes to be left null in root contexts, not by stipulating these contexts but because these null allomorphs are special cases of ungoverned categories.

3.2 Optionally null grammatical elements in governed contexts: still subject to the ECP

We have seen above that governed categories are subject to the ECP, i.e., they must be properly governed; cf. (14). A necessary condition on at least some of these empty categories is apparently that (a movement chain terminating in) their governor X^0 be lexicalized, which then must contribute to proper government. The following contrasts exemplify this in four distinct contexts, with governors in bold:

(26) a. The unmarked C *that* can be null only when its governing V is overt (Stowell, 1985):
John [ᵥ **said**] [c that / ∅] he would quit and Mary [ᵥ ∅] [c that / *∅] he shouldn't.

b. A VP can be ellipted under identity when its governing I is overt and finite (Lobeck, 1986):
He was [ᵥₚ being nice] yesterday, and she [ᵢ **is**] [ᵥₚ being nice / ∅] today.
He was [ᵥₚ being nice] yesterday, and she's [ᵢ ∅] [ᵥₚ being nice / *∅] today.

c. A VP can be ellipted under identity when its governing I is overt and non-finite:[13]
He slowly began to fix it, but it seems that [ᵢ{ **not** / *∅ }] to [ᵥₚ

d. A highest subject in a relative clause can be null only if its governing C is overt:[14]
The boy [cₚ { **that** / *∅ } [ɪₚ [ᴅₚ ∅] left this morning]] was happy.
Cf. The town [cₚ { **that** / ∅ } [ɪₚ we left [ᴅₚ ∅] this morning]] was prosperous.

---

[12]Consequently, these null allomorphs never count as "more economical" than their overt counterparts by virtue of lexical insertion (merge) not having taken place. Rather, from a strictly syntactic point of view, the shorter alternates are simply optional, i.e., *(Have you) been working yourself too hard lately? (The fact) that John smokes a lot Mary finds noteworthy; It is not unusual (for one) to take oneself too seriously.*

[13]In the view of Lobeck (1986), *not* in non-finite clauses can be in I, similar to *n't* in finite clauses. On the other hand, if *to* is an element in I, it is absent at Spell Out. Cf Section 4.2 for an entry for *to* that expresses this.

[14]*Why* proper government by C is needed here is not clear, since I is a closer governor of the empty subject. This paradigm seems to still await a fully satisfactory analysis.

These paradigms strongly suggest that (bold) governors that are themselves null cannot properly govern.

Assuming then proper government is at work in (26), the examples show that a *properly governed* empty category must be more stringently identified than the *un*governed empty categories in (8)–(13), (17) and (19). Properly governed empty categories are either traces as in classic government and binding paradigms, entirely contentless as in (26a), or "deletions under identity" as in (26b–d). Thus it appears that in order to satisfy the ECP, *any interpreted features of an empty category must be present (identified) elsewhere in the string*. Restated:

(27) *ECP Corollary. Properly governed empty categories cannot retain an unidentified feature in LF.*

In contrast, the *un*governed empty categories in (8)–(13) have autonomous content; even the I in (12) indicates "irrealis" and the D in (13) is related to "factivity." Similarly, the ANIMATE content of ungoverned arbitrary PRO subjects in (17) and (19) need not be "identified" elsewhere in the string.

The only example in (26) that involves a lexical morpheme X^0 rather than a co-indexed phrase is (26a), which concerns the possibility of leaving unpronounced the English C *that*. In terms of using parentheses for optional phonological expression, the lexical entry for *that* should be (28), where the notation $\langle$ I $\rangle$ replaces the older subcategorization frame +__IP.

(28) First version: (that), C, $\langle$ I $\rangle$[15]

There are of course contexts where *that* is not optional. We observed above that governed categories can be empty (properly governed) only when their governor is a lexicalized X^0.[16] For example, since the D governor of a topicalized clause as in (13) is empty, an introductory *that* must be overt. Similarly, when a clause moves out of a complement configuration (so that its deep governor no longer governs it), the head *that* must again be present:

(29) John persuaded Mary, they tell me, *(that) she would easily get the job.
     What we explained to her was *(that) her children should stay outside.

The parentheses notation thus has ample uses in phonological lexical specifications π for expressing null allomorphs that alternate with otherwise realized morphemes. The null symbol $\emptyset$ can also be used for π to express non-alternating null morphemes as in (7).

We also seem to have derived formally a somewhat different and yet not implausible conclusion: by the ECP Corollary (27), the unmarked complementizer *that* / $\emptyset$ generated by (28) lacks any interpreted feature, even though it of course has a syntactic label. In particular, its categorial feature C (which I take to be a P in the context __IP) shouldn't be present at the LF interface. We thus still need to formally express the idea that the syntactic category C of *that* is not interpreted.

---

[15]We can note that in the properly governed contexts where *that* is optional, the variants with and without *that* must be equivalent for Economy of Derivation, probably because they satisfy subcategorizations with the same structures and the same number of free morpheme lexical look-ups. That is, an optional null morpheme generated by a lexical specification (π) counts just like an overt one for Economy.

[16]A head-to-head movement chain whose tail is a trace and whose head is lexicalized appears to qualify as a properly governing head, probably as a consequence of Baker's (1988) Government Transparency Corollary.

4.   Parentheses and ∅ in the Inherent Syntactic Part of Lexical Entries

4.1   Justifying the New Notations

Since *that* has an empty allomorph in properly governed positions, by the ECP Corollary (27) its category C must be absent at LF. Suppose then, as suggested in Section 1, that the presence of ∅ among the inherent syntactic features σ of a lexical entry α indicates that the syntactic category of α is uninterpreted at the LF interface.[17]

(30) *LF Category Deletion. In a lexical entry* $\alpha = \pi$, X^0, ∅, F_i, *the symbol* ∅ *means that* X^0 *is not part of the LF of* α.

We thus arrive at the following revision of (28) for the unmarked English complementizer:

(31) Final version: (that), C, ∅, ⟨ I ⟩

Of course, it is at first glance rather strange to say that a morpheme loses its category at LF. But this conception will allow a range of lexical and grammatical phenomena to be expressed rather elegantly. One such consequence has to do with how to express the simple property, "takes a finite complement."

One principal hypothesis of Emonds (2000) is that certain grammatical lexical items are inserted *only* in the PF component: those that are fully specified with uninterpreted features, namely (i) contextual features, (ii) absence of content features (here indicated by the pair X^0, ∅), and (iii) what I introduce there as "alternatively realized" features (mentioned here only in passing) typical of e.g. inflection.

(32) *Late Insertion. (Only) Lexical items that have no interpretable features are inserted only in PF.*

Recall that the notation ⟨α⟩ means, "has a sister phrase whose highest lexicalized head is of category α." Since Late Insertion implies that *that* is absent except in PF (because its only features are contextual and absence of content), *that*-clauses must be selected therefore in syntax or LF by the feature ⟨FINITE⟩ or perhaps simply ⟨I⟩, rather than by mentioning a particular subordinating conjunction *that* in each selecting item. This significantly simplifies the notation for clausal selection.[18]

Introducing ∅ as a lexical syntactic feature should of course have consequences in treatments of categories other than C. Since items generally taken as C are arguably special cases of P that lack interpretable features (Emonds, 1985, Ch. 7), we might expect ∅ to appear with other P in which a clause-selecting context feature ⟨I⟩ is replaced by a transitivity feature ⟨D⟩. In fact the "empty" P *of* in English is just such a case. Its lexical entry is (33):

---

[17]In Emonds (2000, Ch. 1), I introduced a number of features called "marked absence of content" features. My proposal here, perhaps too restrictive, is that ∅, which negates the LF contribution of the main syntactic category of an item, is the only such lexical feature in Universal Grammar.

[18]The question arises, as pointed out to me by both M. Saito and L. Veselovská, as to what makes a selected finite IP project to CP at all, i.e., could *that* perhaps be "inside" IP? While this possibility is intriguing, there is still some tentative motivation (Emonds, 2000, Ch. 1) for why English embedded finite clauses always seem to be CPs.

Although I have not worked out an analysis in detail, the fact that *that*-clauses and *for*-clauses are the only CPs with uninterpreted head Ps in LF can be the basis for explaining why, unlike temporal and WH-headed subordinate clauses, they are not islands. Cf. Emonds (1985, Ch. 7).

(33) of, P, $\emptyset$, (SOURCE), $\langle$ {D / N} $\rangle$[19]

To see how (33) functions, let's first look at a tentative entry for a more typical P *from* (*from the house, from behind the bar, from out there*):

(34) First version: from, P, SOURCE, $\langle$ D / P $\rangle$

The vast majority of members of P are interpreted in LF as expressing spatial or temporal PLACE and/or PATH; *from* rather than *of* is typical. Since we do not want to proliferate symbols, and syntactic categories should preferably be a subset of cognitive categories, the categories P and PLACE should not be independent categories but at least as a first approximation should be identified. Thus, entries for verbs like *put, place, dash, glance, lurk, reside,* etc., which appear with locational PPs of various sorts, need to contain the context symbol $\langle$P$\rangle$ with perhaps some added feature specifications for subclasses of P.

However, each entry should not need to explicitly exclude PPs introduced by the non-locational *of* (or for that matter by other non-locational Ps like *that, if, due to, except, despite,* etc.). We want lexical formatting to reflect this, i.e., once a child knows that *at, in, near, behind, out, off, until, after,* etc. are Ps, it automatically knows they specify PLACE in space or time. But once the child learns specially that the P *of* does not, i.e., that its lexical specification contains P, $\emptyset$, it is unnecessary to stipulate anything further to prevent it from satisfying contextual features such as $\langle$P$\rangle$: *Mary dashed* { *onto / (away) from / behind / *of* } *the bus; Jim put the books* { *onto / away from / behind / *of* } *the desk.*

Thus, if an *of*-phrase is generated in the position of a PP, its lexical head P^0 is eliminated in LF by (30). Hence, an *of*-phrase cannot satisfy a requirement for a PP. Rather, it may satisfy a context requiring __DP if no other case-assigner is present, as in (35a). Alternatively if SOURCE is chosen in (33), *of* can appear in a context __SOURCE^DP (35b).

(35) a. Non-case-assigner requiring $\langle$ D $\rangle$:
Mary was { considerate / a friend } { of / *from / *$\emptyset$ } [DP her neighbor].
We have a lot { of / *from / *$\emptyset$ } [NP different brands].

b. Predicates or other constructions requiring $\langle$ SOURCE $\rangle$:
John { asked / demanded / expected / required } that favor { from / of } me.
Mary is {from / of } the same background.

Let us linger here for a moment to revise the entry for *from*. The feature SOURCE on *from* is obligatory. Consequently, *from* is excluded from (35a) by Economy—it would introduce a superfluous, uninterpretable feature into a context that requires only D. In contrast, the examples (35b) show that the LF interpretation of PATH / PLACE for the feature P on *from* is optional, even though SOURCE is obligatory. If P were interpreted in LF in (35b), we would also expect other Ps at least of SOURCE, contrary to fact: *John expected that favor out of*

---

[19] A small but welcome result of this entry is that the P *of* introducing the indirect object of verbs such as *ask, demand, expect,* etc. is inserted only in PF. In Emonds (1993), I claim that English bare indirect objects as in (i) are invariably derived from PPs whose P is empty at Spell Out, although of course the verbs must satisfy a well known additional criterion, that they consist phonologically of a single initially stressed foot. Since *ask* has this form, its *of*-phrase can also be the basis of a bare indirect object (ii):

(i) Mary { sent / fixed / *delivered / *improvised } her friend a meal.

(ii) Mary { asked / *demanded } her friend a favor.

*me; *Mary is off the same background.* But these constructions are not compatible with such PPs, but only with the feature SOURCE.

When a category is interpreted in LF optionally in this way, I propose using parentheses for it:

(36) Final version: from, (P), SOURCE, ⟨ D / P ⟩

Actually, the category P of the unmarked preposition *to* plays a similar ambiguous role in LF and can be represented almost identically with its syntactic category parenthesized:

(37) to, (P), ⟨ D / P ⟩

I now show in some detail how this entry (37) interacts with the contextual frames for English verbs (transitive or not) subcategorized for an indirect object. Let us restrict attention to transitive verbs whose PF has a single initially stressed foot, as only this class productively allows bare indirect objects. With some of these verbs (38), a *to*-phrase plays the role of a PP and seems unrelated to bare indirect objects. These verbs have a frame based on ⟨P⟩ (with perhaps some other feature), and the P *to* is simply interpreted in LF as a PATH.

(38) He { dragged / lifted / carted } a chair {{ to / toward / away from } Mary / down / out the door}.
  *He { dragged / lifted / carted } Mary a chair.

Other verbs permit a *to*-phrase to alternate with a bare indirect object. Some of these also take PP complements (39a) and some do not (39b).

(39) a. He { handed / took / kicked } a chair {{ to / toward / away from } Mary / down / out the door}.
  He { handed / took / kicked } Mary a chair.

  b. He { read / paid / gave } a bill {{ to / *toward / *away from } Mary / *down / *out the door }.
  He { read / paid / gave } Mary a bill.

As in the analysis in Emonds (1993), the possibility of a bare indirect object signals a PP whose head is absent at Spell Out. The DP within this PP nonetheless has a case feature, identified in that analysis with P itself spelled out on a DP as a secondary, "alternatively realized" feature. That is, such indirect objects of transitive verbs are generated by a contextual frame ⟨D⟩, ⟨[D, P]⟩ in which D rather than P is the primary feature of the second complement.[20]

We have now seen that the lexical entries of four of the least marked members of the English system of Ps, namely *that, of, to* and *from*, utilize parentheses and/or the null notation ∅ in parts of their phonological and/or syntactic specifications. So we have now introduced and motivated two new lexical notations, parentheses and ∅, in the inherent syntactic feature specifications of lexical entries. The rest of this section further justifies these notational innovations.

---

[20]Depending on the selecting verb, one or both of the selected objects may be optional (parenthesized).

4.2   Further Motivations for the LF Cancellation Feature ∅

It was mentioned in Section 4.1 that finite clauses can be selected very simply by the contextual specification ⟨I⟩, since their introductory C (*that*) is deleted at LF. However, infinitives should not be automatically included in selection of ⟨I⟩. Many English verbs take finite complements but exclude infinitival complements (40a), and many others permit infinitives but exclude finite complements (40b):

(40)   a.   complain, conclude, deny, doubt, exclaim, explain, quip, reason, recognize, respond, reveal

      b.   assist, attempt, begin, continue, encourage, force, hasten, help, refuse, try, undertake, yearn

We therefore need to select especially infinitives of obligatory control (with null PRO subjects) by some contextual specification other than ⟨I⟩. In this vein, Jackendoff (1975) shows that the semantics of the complement structures "V+(DP)+finite clause" consistently differs from that of "V+(DP)+obligatory control infinitive." While the former involve the semantics of propositions, thereby plausibly reflecting selection of I, the latter are closely related to the semantics of events, suggesting selection of V.

It seems then that the infinitival marker *to* (and probably its bound morpheme counterpart in many languages) is extraneous to how infinitives are lexically selected. That is, obligatory control infinitives should be selected simply by the context feature ⟨V⟩ (" has a sister phrase whose highest lexicalized head is of category V"), even though these VPs may further project to IP to satisfy the requirement that Vs have subjects—the Extended Projection Principle of Chomsky (1981, Ch. 2).[21] The head I of this IP cannot then be lexicalized when VP is selected. That is, this I is without content and is lexicalized only in PF by (32). The item in the English lexicon satisfying this description is the marker *to*, whose category feature disappears in LF by means of the cancellation feature ∅:

(41)   to, I, ∅[22]

Emonds (2000, Chs. 7 and 10) argues in some detail that late insertion of *to* predicts contexts of obligatory PRO without invoking the "PRO theorem" of government and binding. Since obligatory control is restricted to complement clauses (a controlled subject PRO alternates with overt subjects and arbitrary PRO in subject and adjunct clauses), it should result from some selectional device; therefore, deriving it from a subcategorization feature ⟨V⟩ correctly predicts its absence in subject and adjunct clauses.

There is an additional complexity concerning the role of LF cancellation. It is observed in Emonds (2000, Ch. 7) that in some obligatory control contexts, infinitives with a more "irrealis" sense (42a) contrast with more "factive" gerunds (42b).

(42)   a   Ann tried to study history.   The boy forgot to study history.
      b   Ann tried studying history.   The boy forgot studying history.

---

[21]I don't exclude VP-internal subject DPs, but if they are lexical then at least in English a VP must project to IP to furnish a lexical DP subject with case. If infinitival VPs do not project to IP, if of course follows that the category feature of *to* is not I but some lower functional head. The reasoning in the text remains the same; the category of the lexical entry of *to* must be invisible at LF so that obligatory control infinitives can be selected by ⟨V⟩.

[22]I assume that the occurrence of VP sisters with all I is not due to lexical selection with each member of I.

To express this formally, I suggest there that these infinitives are selected by a contextual feature ⟨[V, M]⟩ rather than simply by ⟨V⟩.[23] This implies that a contextual feature can unite two head categories, where the complement's lexicalized head is indicated by the *initial* category X (= V for this example). The second category Y (here M), which cannot of course actually appear on V, must be generated on an empty head *between* the selecting item α and X. Hence, the feature ⟨ [V, M] ⟩ for a verb like *try* selects the structure (43).

(43)
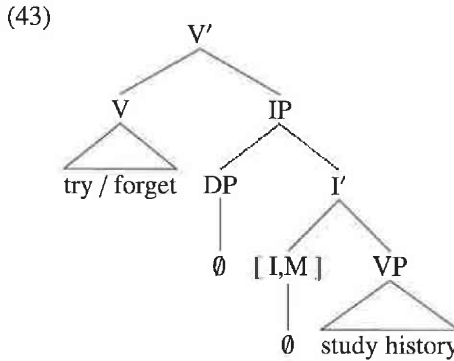
Because of the feature ∅ in the entry (41), the syntactic category I of *to* deletes in LF. M remains and provides the less "realis" sense in LF that differentiates the infinitives from the gerunds in (42). On the PF branch, by (32), *to* is inserted under I.[24] Since a featureless I does not assign nominative case at Spell Out, the subject DP in (43) must be empty, giving rise to "obligatory control."

A final grammatical item that we can mention as containing the lexical specification ∅ is the copula verb *be*. The minimal and unmarked meaning of the category V in LF is not currently clear; possible candidates are ACTIVITY (= −STATIVE), CHANGE OF STATE and Accusative Case Assignment. The first two of these features are respectively spelled out by the grammatical verbs *do* and *get* in English. But in any event, most and perhaps all uses of *be* clearly have none of these meanings. That is, whatever feature or feature complex V may generally realize at LF, it appears that V should be cancelled for the correct interpretation of *be*, i.e., the feature pair V, ∅ should occur in its lexical entry. Schütze (2001) provides an extensive defense of *be* as a null V at LF.

4.3   Further Motivations for Parenthesizing Syntactic Categories

We have not yet seen uses for the new lexical notations for elements in the categories D, N or A. We can begin with a rather obvious case where the syntactic category D of certain morphemes is not interpreted at LF—the case of expletives. There would seem to be little doubt that the personal pronouns including *it* are instances of D. They are invariably definite, and many of them occur in positions of definite articles (Postal, 1969): *we three young doctors, you old fool(s), them nice girls* (non-standard usage).

A further characteristic of some of these pronouns (third person pronouns, *who*, *I*) is that they are incompatible with the overt NPs that usually accompany D. This marked situation,

---

[23]The analysis referred to shows why the simple context specification ⟨V⟩ for these verbs selects gerunds with accusative case in direct object position rather than caseless infinitives.

[24]Conceivably, one might need to revise (41) slightly to accommodate this analysis, as follows: *to*, I, (M), ∅.

complete absence of a sister phrase at both PF and LF, is expressed by the context specification $\langle \emptyset \rangle$ introduced in Section 2.

In addition, for expletives such as *it* we need to express the fact that they do not express reference.

(44) it, (D), $\langle \emptyset \rangle$[25]

The parentheses in (44) mean that any (projection of) D is optionally eliminated in LF representations of *it*. Of course, ill-formedness results if this happens in any kind of argument position, since arguments must be DPs in LF (Longobardi, 1994). But by the same logic, D *must* be eliminated in any non-argument position, and (44) makes this possible for *it* (but not for other pronouns).

A range of data suggests that *there* is also a pronoun, marked both for LOCATION and Definiteness (D). In (45a), *there* satisfies the subcategorization of *from* for DP and also serves as an antecedent for a pronominal DP. In (45b), pronominal *there* satisfies contexts that require LOCATION. The "Definiteness Effect" in the predicate nominal (45c) shows that *there* is Definite.

(45) a. Mary came from there$_i$, even though she thought it$_i$ was a boring town.

      b. Bill { dashed / lurked / put the books } there a few minutes ago.

      c. There must be { somewhere / someplace / *that place / *there } that I can live.

Since *there* like *it* can also appear in nonargument positions, in particular in certain surface subject positions such as the initial position in (45c), its lexical entry must be closely akin to (44), with its syntactic category in parentheses:

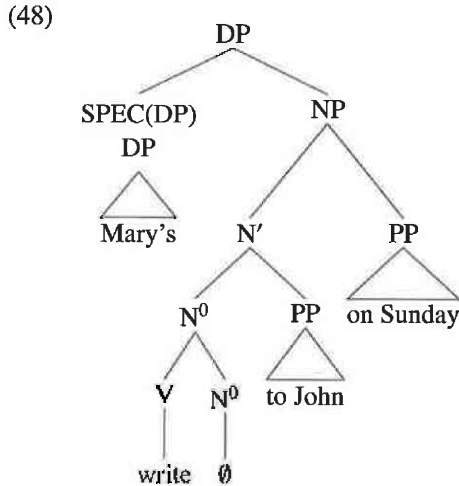(46) there, (D), LOCATION, $\langle \emptyset \rangle$

We have now seen several instances of free morphemes that can or do lose their category X at LF via the inherent lexical specifications (X) or $\emptyset$: *that, of, from, to* (P), *to* (I), *be, it, there*. These formatting devices can also function with bound morphemes, especially those traditionally thought to express related but yet distinct inflectional and derivational processes.

For example, consider the English noun- and adjective-forming suffix on verbs *–ing*. The classical study of Chomsky (1970) clearly distinguishes "derived nominals" and "mixed nominals," which have a range of nominal properties as in (47a), from "gerundive nominals," which act like they have a verbal head (47b). As is well known mixing the properties leads to ungrammaticality as in (47c).

(47) a. Some of Mary's imprudent writing of notes to John (determiners, adjs, *of*-phrases)

      b. Mary's having imprudently written notes to John (aspect, adverbs, direct objects)

      c. *Mary's imprudently writing of notes upset us.
         *Mary's imprudent writing notes upset us.
         *Some of having written notes was imprudent.

---

[25] In the absence of any compelling contrary reason, I assume that a D unaccompanied by any marked feature of quantification or numeric value simply means "Definite" in LF. If not, a feature DEF can be added to (44). Lexical entries presumably do not explicitly list unmarked features, such as –PLURAL and III PERSON of *it*.

In Emonds (1991; 2000, Ch. 4), I argue that in fact taken abstractly, both types of nominals have the same structure, exemplified in (48) for the noun phrase *Mary's writing to John on Sunday*. What differentiates the two types of nominalization is not structure but the derivational level at which *–ing* substitutes for ∅, the "right hand head" of the larger N^0.

(48)

```
                          DP
                    ┌──────┴──────┐
              SPEC(DP)            NP
                DP          ┌──────┴──────┐
                △          N'            PP
             Mary's   ┌────┴────┐       △
                     N^0       PP    on Sunday
                  ┌───┴───┐    △
                  V      N^0  to John
                  │       │
                write     ∅
```

Suppose that *–ing* has a lexical entry as in (49):

(49) ing, ({ N / A }), ⟨ V__ ⟩[26]

If the parentheses around the category feature(s) {N / A} in (49) are ignored, then *–ing* is inserted normally (prior to Spell Out) and its category feature N (or A) is interpreted in LF. Supposing it is N as in the examples under discussion, this N is capable of being assigned reference by elements in DP when NP projects to DP, as is typical. Its head noun *writing* then governs like an N: it can't assign accusative case, its modifiers are adjectival in form, it can often be pluralized or quantified, and it exhibits any of the properties of derived or mixed nominals noted in Chomsky (1970).

If the parentheses around the category feature(s) in (49) are on the other hand "chosen," then this category (N) is cancelled (i.e., deleted) at LF. Since this entry has no other interpretable feature, Late Insertion (32) guarantees that *–ing* enters the tree only at PF. Consequently, the lexicalized head of (48) during the syntactic derivation is V, which then predictably assigns accusative case to DP objects, requires modifiers that are adverbial in form, is compatible with aspectual auxiliaries, and in general exhibits the properties of gerundive nominals noted in Chomsky (1970).

Emonds (2000, Ch.5)—cf. also the preceding issue of this journal—provides an analysis of the ambiguous English passive morpheme *–en*, whose theoretical aspects are similar to the above analysis for *–ing*. That discussion shows how the derivational use of [_A *en*] corresponds to passive adjectives, which clearly exhibit the characteristic LF "property interpretation" of the category A, while its inflectional use corresponds to "verbal passives," which are so-called

---

[26]The disjunction { N / A } simply means that *–ing* can form nominals, either derived or gerundive as in (47a–b), *or* adjectives, again either derived (*very interesting, so boring*) or participial. In most treatments, this disjunction is simply taken as a shared cross-categorial feature written as [+N].

precisely because this construction acts and is interpreted like its head is a V. Hence the lexical for entry for *–en* is as in (50), where F_i is a grammatical feature particular to passives related to "absorption" of the direct object:

(50) en, (A), ⟨ V__ ⟩, F_i

A final example of the parenthesis notation for inherent syntactic categories is furnished by the Japanese passive suffix *–(r)are*. In a detailed analysis of Japanese passive structures, Kubo (1992) revises a traditional division into direct passives (those translatable into English) and indirect passives (those not so translatable). She finds that the linguistically significant division is rather between (i) passive predications which *must* (not may) be adversative for the surface subject and arguably never exhibit a gap corresponding to a moved DP ("gapless passives") and (ii) those which need not be adversative and arguably always contain a trace ("gapped passives").[27] Her analysis attributes gapless passives to the fact that [$_V$ *–(r)are*] may assign a theta role to an external argument (thereby blocking movement), and gapped passives to the fact that it equally well can fail to do so, which in turn forces NP Movement of some argument into the surface subject position so as to fulfill the EPP requirement that the V *–(r)are* have a surface subject.

Although Kubo concludes her study with a succinct and elegant lexical entry for *–(r)are*, it crucially uses the difference between insertions in the contexts VP__ for type (i) and V__ for type (ii). The system of lexical notation adopted here excludes such notation. Instead, it seems to me that the simple entry (51) suffices. *–(r)are* selects a V-headed phrase; then like several other Japanese grammatical V and A (e.g., *-tai* 'want to'; *-(s)ase*, 'make'), it agglutinates with its V complement at PF.

(51) –(r)are, (V), ⟨ V ⟩

When the category V in (51) is retained at LF—i.e. its parentheses are ignored—it must independently exhibit a minimal V-like semantic property. This consists of assigning a theta-role to its deep subject, possibly by a stipulation added to (51) but also conceivably by some default mechanism.[28] Consequently, movement to that subject position is impossible. On the other hand, when V in (51) is cancelled at LF—i.e., its parentheses are chosen—nothing remains to interpret with *–(r)are* and it is inserted in PF. Its subject is then empty and, as Kubo argues, a DP must raise to SPEC(IP) from any of a variety of syntactic positions to satisfy the EPP (or whatever mechanism is taken to explain the EPP).

We can conclude that the parenthesis notation for inherent syntactic categories, implying optional cancellation of the LF value of those categories, is well motivated by Japanese passives as well as by the versatile English bound morphemes *–ing* and *–en*.[29] When parentheses

---

[27] Although *–(r)are* does not assign case separately from the V it is suffixed to in Kubo's system, as a V (as seen by its own tense morphology) it does not prevent its host V from assigning case either. Consequently, she claims that the sole motor for movement in Japanese passives is attraction of some DP to the unfilled subject position. (She further argues that Japanese lacks expletive subjects.) Her analysis thus prefigures the approach to passive movement in Chomsky (1995).

[28] Kubo's entry for *–(r)are* stipulates optional assignment of a malefactive theta role. My tendency is to think that a non-activity verb can assign a benefactive / malefactive role as a default, with any limitation to positive or negative connotation being added by an inherent "evaluative" feature on the V.

[29] We have not gone through examples for how the parallel alternation between derived adjectives and present participles works when *–ing* is an A; paradigms and arguments are given in the sources cited.

are ignored, the suffix's category is interpreted and what is called a derivational bound morpheme is inserted in syntax. When they are chosen, the same category is uninterpreted (the stem's category remains syntactically active in a derivation), yielding what is called an inflection inserted in PF.

At the same time, the fact is that most bound morphemes do *not* cross the inflection / derivation border in parallel ways. There is probably a modest lexical cost associated with "generalizing" a derivational morpheme with parentheses as in (49)–(51), but one that is less than introducing a novel entry for a new inflection. Moreover, some morphemes that are purely inflectional may well have the cancellation feature $\emptyset$ in their lexical entry, since it is improbable for example that a number agreement suffix on verbs is interpreted as a separate V in LF.

5. Conclusions

The introductory section observed that current grammatical theorizing claims that syntactic structures are based on properties of lexical entries, and yet has essentially accomplished next to nothing in terms of contentfully specifying what the latter might be, i.e. of characterizing "possible lexical entry." What results is an apparently formal theory about entirely informal and intuitive objects. In my view, a central reason for this paradoxical state of affairs is a lack of concern for actually stating the lexical entries. We need to get beyond truisms, stated with or without Greek letters, about "sets of phonological, syntactic and semantic features." Especially with respect to closed class items, we need to set as a goal finding clear notions and making nontrivial claims about which symbols and algebra of symbols are, might be, or probably are not possible in formalized lexical entries.

To this end, I feel that Emonds (2000) has proposed some plausible and important restrictions on what categories appear in lexical entries: (i) the claim that syntactic features all have their basis in cognition;[30] (ii) the exclusion of any purely semantic combinatory features; (iii) the extension and unification of syntactic subcategorization in both word-internal and phrasal syntax; (iv) the restriction of all lexical statements to non-phrasal categories.

This paper has moved on to specifying some combinatory possibilities for these categories. That is, it proposes some non-trivial clarifications for an algebra that combines the categories and symbols in lexical entries. In particular, I have argued for extending use of both the parenthesis notation and the symbol $\emptyset$ meaning "null or absent at an interface" to all the non-purely semantic parts of lexical entries. Which interface $\emptyset$ refers to depends on where in an entry $\emptyset$ occurs, as motivated in the paper's various sections.

My conclusions can be summarized formally as follows. Suppose as in the introductory section we have a lexical entry $\alpha = \pi, \sigma, \langle \gamma \rangle, \lambda$, where γ is a set of syntactic contextual features σ'. The following conventions of Universal Grammar govern the interpretation of the symbols $\emptyset$ and parentheses in lexical entries.

(52)   Parentheses and $\emptyset$ in the phonological part of lexical entries:

---

[30]That is, there are no "purely formal features" of syntax not ultimately drawn from the set of semantic features Λ. Although case features are often talked about as if they were not, I have argued in Emonds (1985, Chs. 1 and 5; 2000, Ch. 7) that these features are "alternative realizations" of the case assigners themselves, e.g., V, P, etc. These case-assigning categories play a central role in LF in representing ACTIVITY, CHANGE OF STATE, LOCATION, etc. To my knowledge, my arguments for these claims about case features have not been refuted (or for that matter, discussed). Other current treatments of case categories remain essentially indistinguishable from nineteenth century Indo-Europeanist taxonomy.

    a. If $\pi = \emptyset$, then α is null at the PF interface.

    b. If π is parenthesized, then it alternates with $\emptyset$ at PF.

(53)   Parentheses and $\emptyset$ in the syntactic part of lexical entries:

    a. If $\sigma = \emptyset$, then α has no syntactic cateogry: *yes, gosh, hey, ouch, hello*, etc.

    b. If $\sigma = (F_1), F_2, \ldots$, then α has two uses at LF, one where F_1 is absent at LF and one where F_1 plays a role.

    c. If $\sigma = F_1, \emptyset, F_2, \ldots$, then F_1 is absent ("cancelled") on α at the LF interface.

(54)   Parentheses and $\emptyset$ in the contextual part of lexical entries:

    a. If $\langle \gamma \rangle = \langle \emptyset \rangle$, then α can have no sister γ at LF or PF.

    b. If $\langle \gamma \rangle = \langle (F_1), F_2, \ldots \rangle$, then α appears with or without a sister γ with the feature F_1 at both LF and PF.

    c. If $\langle \gamma \rangle = \langle F_1, \emptyset, F_2, \ldots \rangle$, then one choice for an obligatory sister γ of α at PF and LF is a "phonetically null discourse anaphor."

Finally, π cannot be $\emptyset$ if α contains any λ separate from σ. This rephrases (6).

References

Baker, Mark. 1988. *Incorporation*. Chicago: University of Chicago Press.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

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

Chomsky, Noam. 1970. Remarks on nominalizations. In his *Studies on Semantics in Generative Grammar*. The Hague: Mouton.

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

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Emonds, Joseph. 1985. *A Unified Theory of Syntactic Categories*. Dordrecht: Foris.

Emonds, Joseph. 1991. The autonomy of the syntactic lexicon and syntax. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, C. Georgopoulos and R. Ishihara, eds. Dordrecht: Kluwer Academic Press.

Emonds, Joseph. 1993. Projecting indirect objects. *The Linguistic Review* 10. 211–262.

Emonds, Joseph. 2000. *Lexicon and Grammar: the English Syntacticon*. Berlin: Mouton de Gruyter.

Emonds, Joseph. 2001. Passive syntactic structures. *Theoretical and Applied Linguistics at Kobe Shoin* 4. 1–29.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10. 279–326.

Hendrick, Randall. 1982. Reduced questions and their theoretical implications. *Language* 58. 800–819.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671.

Kubo, Miori. 1992. Japanese passives. *Hokkaido University Institute of Language and Culture Studies* 23. 231–301.

Lieber, Rochelle. 1980. *On the Organization of the Lexicon.* Doctoral dissertation. Massachusetts Institute of Technology, Cambridge.

Lobeck, Anne. 1986. *Syntactic Constraints on VP Ellipsis.* Doctoral dissertation. University of Washington, Seattle.

Lobeck, Anne. 1995. *Ellipsis: Functional Heads, Licensing, and Identification.* Oxford: Oxford University Press.

Longobardi, Giuseppe. 1994. Reference and proper names: a theory of n-movement in syntax and logical form.' *Linguistic Inquiry* 25. 609–655.

Postal, Paul. 1969. On so-called "pronouns" in English. In *Modern Studies in English: Readings in Transformational Grammar,* D. Reibel and S. Schane, eds. Englewood Cliffs: Prentice-Hall.

Schütze, Carson. 2001. Semantically empty lexical heads as a last resort. In *Semi-lexical Categories*, N. Corver and H. van Riemsdijk, eds. Berlin: Mouton de Gruyter.

Stowell, Timothy. 1985. Null operators and proper government. *Proceedings of NELS 16.* Amherst: University of Massachusetts Graduate Linguistic Student Association.

Zagona, Karen. 1982. *Government and Proper Government of Verbal Projections.* Doctoral dissertation. University of Washington, Seattle.

Author's email address: jeemonds@sils.shoin.ac.jp