# LINKING LANGUAGE-PARTICULAR MORPHOLOGY WITH UNIVERSAL SYNTAX[1]

Joseph E. Emonds
Kobe-Shoin Women's University

This essay proposes to sketch some consequences of the idea that language-particular lexical entries for grammatical morphemes (i) really do account for the many detailed morphological and grammatical differences among languages, and (ii) are actually the *only* language-specific properties that we need postulate among languages in order to account for these differences. Otherwise, with the possible exception of word order parameters, the derivational syntax and resulting logical representations of the "content" of sentences are general and cross-linguistically uniform.[2]

I approach this problem with a derivational generative model that, at least broadly, is "minimalist" as in Chomsky 1995. In particular, derivation proceeds in "phases" from bottom to top in a tree: in a given "phase" (previously, roughly "cyclic domain"), content items from a lexicon are assembled in trees according to well-formedness conditions and then transformed in a derivational computation. The phrasal structure so derived can be a unit within a possible larger domain in which the cycle of operations can repeat.[3]

## 1. Uniform Representations of Universal Syntax

I assume that in order to compare natural languages formally, say in scientific models or computation, some level of syntactic representation must be both real, i.e.,

---

[1] This title can also describe the efforts of individual researchers to link fields of language-particular studies, often in fact morphology-based, to late twentieth century concerns of universal syntax. It is Rudolph de Rijk's singular and outstanding contribution to have forged this link for the unique human treasure of the Basque language. The fact that this field is now thriving is a testament to the success of his lifelong mission.

[2] This paper revises an earlier contribution in *Newcastle & Durham Working Papers in Linguistics*, vol. 3, published jointly by the Department of Linguistics and English Language at the University of Durham and the Centre for Research in Linguistics at the University of Newcastle upon Tyne. I am grateful to the heads of those units, Professors Anders Holmberg and Noel Burton-Roberts, for permissions to use that text as a basis for this one. I wish to thank Lida Veselovska for careful reading and suggestions for improvement of that paper. Remaining inadequacies are of course my own.

[3] During a given phase, the heads and (some) edges of any already processed cyclic domains may be accessed and further modified before the structure of the phase in question is interpreted as paired phonological and semantic (or "logical") forms. Another type of partial access to a processed domain is what I call "late lexical insertion". Emonds 2000 argues that many syntactic generalizations can be properly expressed only if a number of grammatical elements (but no content elements) are inserted in trees *after* transformational computation on a domain.

[ASJU Geh 44, 2002, 179-198]
http://www.ehu.es/ojs/index.php/asju

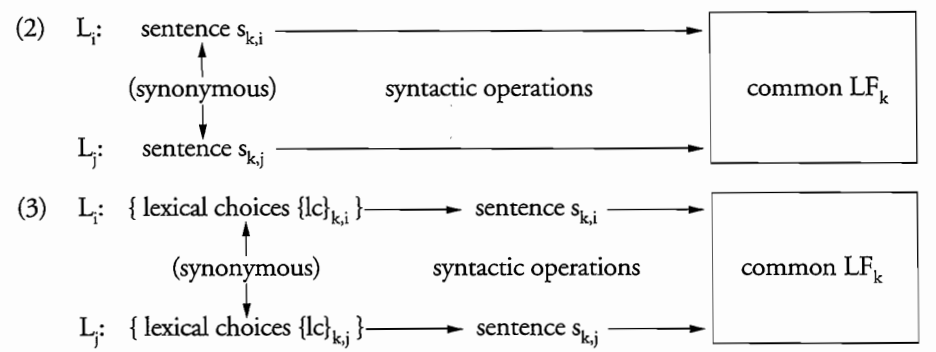
actually reflect properties of occurring natural languages, and at the same time universal. Notice that a claim that different languages share universal representations goes beyond a more elementary requirement that descriptions of individual languages simply conform to some restrictive principles of combination, as in earlier generative grammar.

Fortunately, in most current Chomskyan models of syntax there seems to be an emerging intuition that a discoverable syntax for natural languages has a fairly close relation to universal representations. The idea is basically this:

(1) Each (unambiguous) sentence has a universal representation of its meaning/ semantics which is the endpoint of a syntactic derivation. We call this the Logical Form of the sentence.

That is, the Logical Forms (LF) of synonymous sentences are typically *the same across languages*, and furthermore result from derivations performed in a syntactic component. What differs across languages is the inputs to these derivations, i.e., the lexical items which comprise individual languages.[4] Of course, the idea that languages share LFs advances the work of actually determining these LFs only if we have independent evidence as to their form drawn from work on particular languages.

The syntactic derivation of a sentence is thus the series of structures that result when a language-particular set of lexical choices is combined by language-independent well-formedness conditions and transformational operations so as to yield the LF of that sentence. All of these conditions and operations then constitute what is called Universal Syntax (or Universal Grammar). For two synonymous sentences, one in one language $L_i$ and another in a different language $L_j$, this can be initially schematised as (2), and more accurately as (3):[5]




[4] Precisely because the lexical choices and hence certain aspects of syntactic derivations differ from language to language, certain combinations leading to a certain $LF_k$ in one language may simply not be available in another, as interestingly argued in Keenan 1975. Therefore, (1) does not imply that each language encompasses exactly the same set of LFs

[5] This diagram reflects an essential aspect of a Chomskyan methodology for discovering LF representations of meaning, and underscores a corresponding weakness of semantics-based methodology. Namely, research starts from some testable (non-speculative) knowledge about the various sentences $s_{k,i}$ in a language, and not with the LFs ("meanings") themselves. In contrast, semantics-based theorising essentially tries to intuit the properties of LFs (e.g., theta roles, quantification structures, lexical conceptual structures) and then circularly tries to derive properties of universal syntax from these guesses.

## 2. The Lexicon as the Source of Language-Particular Syntax

Typically, the sentence structures $s_{k,i}$ and $s_{k,j}$ from different languages $L_i$ and $L_j$ in (3) differ in ways that go well beyond obvious differences in word order and the phonology of lexical choices. Nonetheless, according to a further restrictive hypothesis first formulated in Borer 1984, the syntactic differences among languages (e.g., between synonymous $s_{k,i}$ and $s_{k,j}$) all result from variations in the lexical specifications of the grammatical morphemes of each language. That is, counter to pedagogical and descriptive linguistic traditions that speak of a particular language's "grammar," generativists of a Chomskyan bent now usually adhere to (4):

(4) Language-particular syntax resides entirely in the inherent and contextual lexical features of closed class lexical items.[6]

It is not news to the practiced grammarian that closed class lexical entries with the "same meaning" (or with "no meaning") can differ across languages. Corresponding entries can be of different categories, be bound or free morphemes, subdivide categories differently (two vs. three demonstratives, finer subdivisions of tenses), occur in distinct grammatical contexts, etc. For example, let us compare the sentential negations in French (an invariant free morpheme) and Japanese (an adjectival verb fully inflected for TENSE values in I; cf. Kato 1985).

The most common way to negate a French finite clause is with this invariant *pas* following (not necessarily adjacent to) the inflectional position I of the verb, i.e. with *pas* initial in the VP: *Marie (ne) serait certainement* $[_{VP}$ ***pas*** *venue si tard]* "Mary would certainly not have come so late."

The Japanese counterpart, in contrast, uses a necessarily preffixed and hence bound morpheme *na-*; moreover, its conjugational suffixes expressing the categories of finite inflection $I^0$ are those which occur only with adjectives (A). Japanese sentential negation is therefore a typical phrase-final head A *na-* (bold in the following example) which agglutinates with both a preceding head of a VP and a following inflection *katta* "Copula + Past": *Taro nara* $[_{VP}$ *sorehodo osoku* $[_V$ *ko]* $]$ ***na****katta daro* "Taro would certainly not have come so late".

Hypothesis (4) attributes such language-particular differences to lexical differences. It suggests that French and Japanese sentential negations result from different lexical entries, which we can formulate as (5):

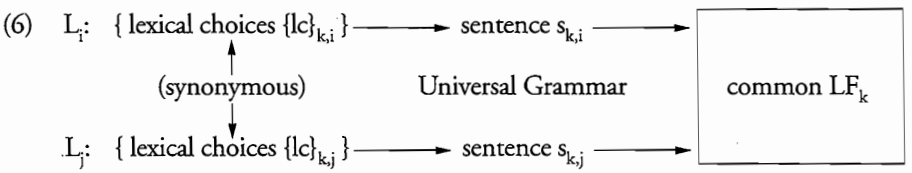
---

[6] One obvious language-particular syntactic property that resists formulation in terms of (4) is the "head-initial" vs. "head-final" parameter implicit in Greenberg 1963 and explicit in Stowell's 1981 formulation. It is by now well-known that some languages present non-obvious variations on the "pure versions" of head ordering; German verb and adjective phrases have at least superficial head-final properties, although its other phrases appear to be head-initial. Chinese verbs appear to precede their complements, while otherwise the language is head-final.

In a Seattle lecture in the late 1980s, N. Chomsky responded to the author's question about reconciling this parameter with (4) by suggesting that word order may be a "global property of the lexicon". That is, left-to-right ordering would result from a lexical contextual feature of heads constant throughout a language —and thereby conform to (4). The idea of lexicons having global but nonetheless language-particular formatting is tantalising, but I have not seen it pursued.

(5) a. French: pas, NEG, { SPEC(VP)/ +___XP in root contexts, ... }[7]
b. Japanese: na, NEG, A, +VP___, +___$I^0$

What is new in Borer's proposal (and succeeding variants) is the idea that *all* of language-particular syntax can be reduced to differences such as seen in (5). This permits us to sharpen the scheme (3) to (6):

(6)



Of course, languages are not somehow miraculously brought closer together by Borer's hypothesis. Language-particular complexities arise quickly by compounding the many permitted lexical differences between corresponding morphemes. To see how quickly, let us combine the differences in French and Japanese sentential negations (5) with differences generated by the same two languages' contrasting passive morphemes.[8]

As is well known, the productive French passive participle form is a bound verbal suffix which shows adjectival agreement in gender and number.

(7) French passive: é, A, +V___, +F

The F in (7) is some additional feature which distinguishes passive from active participial and other adjectival endings such as *-ant* "-ing" in *intéressant* "interesting", *épatant* "surprising", etc.; we return to the nature of F at the end of section 4.

The Japanese passive morpheme *-(r)are* is in sharp contrast to French *-é*. Though also a verbal suffix, it takes specifically verbal, not adjectival inflections, and thus also contrasts with the Japanese negative *na-* in (5). Its verbal character is a point of departure for Kubo's 1992 study of passives, and her study culminates with a lexical entry for *-(r)are* which we can adapt here.

As in most studies of Japanese passives, she finds that two somewhat different constructions go by this name. First, when *-(r)are* assigns its own theta role (malefactive) to an underlying subject, it heads its own clause and has a $V^{max}$ complement which is "gapless" (all its arguments are overt):

---

[7] Besides its standard use as sentential negation in SPEC(VP), *pas* has other functions such as negating a non-clausal XP in isolation: *pas nous* "not us", *pas ce soir* "not this evening", *pas si vite* "not so fast", *pas avec un type comme lui* "not with a guy like him", etc.

[8] A moment's reflection about the model (6), in which all syntactic structures «project» from the lexicon, indicates that fully determining and precisely formatting the lexical properties of entries such as (5), (7) and (8) must be a basic research concern. Central to this effort would be comparison of elements that roughly correspond cross-linguistically, such as for example copular verbs. For example, "comparative grammar" of e.g. English and Japanese verbal systems cannot rest on merely an impressionistic grasp of the differences between English copulas (forms of *be* and *have*) and those of Japanese (*iru, aru, da, desu, na*). Unfortunately, much syntactic research of the past twenty years treats lexical specifications as matters of at most passing interest required only for expository purposes or «to fix ideas». But grammatical analysis and theorising based on (6) cannot really lay claim to being generative grammar in the term's original sense if the structures in syntactic derivations are «projections» of nothing more than linguists' unformalised intuitions.

(8) a. Japanese gapless passive: (r)are, V, +VP___, +___$I^0$, malefactive[9]

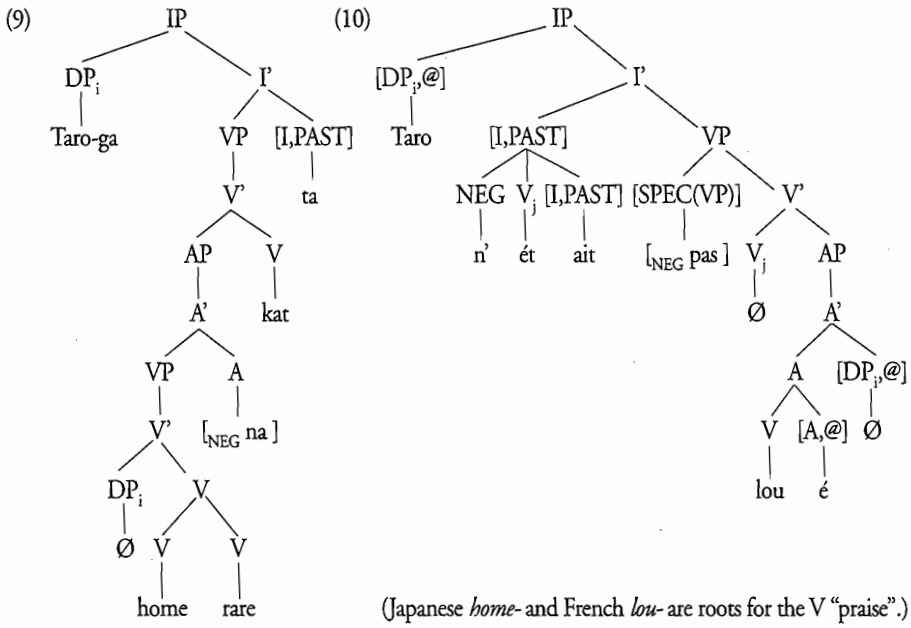
When *-(r)are* assigns no theta role, the subject NP is underlyingly empty, which leads to NP preposing and hence passive VPs which contain gaps. She proposes that Economy of Representation then requires that this contentless *-(r)are* be a simple suffix on a sister $V^0$ inside a single VP.[10]

(8) b. Japanese gapped passive: (r)are, V, +$V^0$___, +___$I^0$

Putting together (8a-b), Kubo obtains (8c) as her final entry for the Japanese passive:

(8) c. Japanese passive: (r)are, V, +$V^k$___, +___$I^0$, (malefactive)

Using (5) and (7)-(8), together with some universal bar notation principles discussed briefly in the next section, we obtain the following two very different trees translating the negated passive sentence "Taro was not praised." For the French tree (10), @ = -FEM, -PLUR, and the highest V is represented as raised to I, a point to which we return.




(Japanese *home-* and French *lou-* are roots for the V "praise".)

Under the model in (4) and (6), these two quite distinct trees, as determined by the differing lexical items for corresponding French and Japanese closed class morphemes, undergo transformational derivations determined by identical principles of UG. Since

---

[9] The French passive participle morpheme has no relation to any separate theta role. Rather, benefactive as well as malefactive relations to an activity are expressed by additional NPs in "affected datives" (Authier and Reed 1992) and "reflexive causatives."

[10] (9) is a typical Japanese pasive with a direct object gap. Examples of other types of gapped passives are given below as (21) – (23).

these two sentences are essentially synonymous, these principles eliminate or make invisible any differences between (9) and (10), transforming both trees into essentially the same language-independent LF (11). Braces replace brackets in (11) to suggest the absence or irrelevance of left-right order in LF, at least as regards grammatical items.[11]

(11) $\{_{IP}\{_{DP}$ Taro $\}_i\{_{I'}\{$ I, PAST $\}, \{_{VP}$ NEG, $\{_{X'}\{_V$ praise $\}, \{_{DP}$ $t_i\}\}\}\}\}$

Presumably, the passive morphemes and any agreement features are among those deleted in the derivations; moreover their categorial projections as well as the syntactic category of NEG (SPEC(VP) vs. A) are irrelevant in LF.

## 3 The Role of Universal Syntax and Language-Particular Word Order

Several aspects of the paired trees in (9)-(10) and the corresponding lexical entries (5) and (7)-(8) are *not* related to lexical differences, and I treat them each in a brief subsection. First, the types of categories, their modes of combination, and even the movement of phrases are basically the same in the two trees and hypothesised as consequences of universal syntax (section 3.1).[12] Second, several differences between the Japanese (9) and the French (10) reduce to the well-known contrast between the syntactic head-initial/ head-final parameter. Although language-particular word order may fall outside lexical variation, improved formulations of the head-initial/ head-final parameter can help us better isolate the residue of language-particular differences due to lexical variation (section 3.2). Third, once word order variation is factored out, the format of the lexical entries across languages turns out to be uniform (section 3.3).

### 3.1. Categorial Uniformity and Hierarchical Universality

In the mid-eighties, generative grammar had seemingly come to a consensus that a restrictive and universal set of syntactic categories called the "bar notation" was empirically adequate and cross-linguistically appropriate. All categories in any language were to be represented as one of four lexical heads (X = N, V, A, P), non-maximal or maximal projections of these heads (respectively X' and XP), and specifier (SPEC) daughter nodes of XP. Some authors further felt that these could be supplemented by a very few functional projections such as D/DP (associated with NP) and I/IP (associated with VP). In line with this emphasis on a parsimonious set of categories, Emonds (1985: Ch. 7) argued that the functional head C for introducing clauses could be considered as a special case of P.

That work's Introduction summarises two implicit and then current principles of universal syntax as follows:

---

[11] Nonetheless, left-right order among phrases may play a role in specifying relations among lexically specific content items in LF, for example in representing the functional sentence perspective of the Prague School.

[12] An alternative for (10) based on Kayne 1991 holds that in underlying structure the French participial ending is a head ADJ with a VP sister. Under this view, the underlying phrasal combinations in (9)-(10), for the negated structures under consideration, become identical. Even so, the two syntactic derivations remain different, but in this case (again) because the closed class items in the two trees have distinct lexical properties.

(12) a. Categorial Uniformity. The categories defined in terms of the bar notation, $X^j$ and SPECIFIER(XP), do not differ from language to language, but their subcategories which are realised in each language's syntax may vary.

b. Hierarchical Universality. The range of permitted hierarchical combinations of syntactic categories does not vary from language to language at the level of deep structure.

Since then, many analysts have replaced syntactic features of the six lexical and functional heads with an expanded set of functional heads. I remain unconvinced of the fruitfulness of this direction and find that a more parsimonious theory retains its appeal and promise. Therefore, the trees in this study use only the four lexical heads (X = N, V, A, P), two functional heads D and I, and specifier (SPEC) daughter nodes of XP, which may carry a limited number of syntactic features such as NEG. In accord with argumentation brought together in Speas (1990: 2.2), an XP is simply an X' which doesn't further project; it is not a type of category separate from X'. In this sense, the French and Japanese constructions being compared in (9)-(10) are alike in that they utilise the same categorical inventories.

### 3.2. Unmarked and Marked Headedness in Syntactic Systems

Variations in left-to-right ordering of grammatical elements contrast with the uniform cross-linguistic category systems (12a) and dominance relations (12b) imposed by the universal bar notation. The orderings found in the trees (9)-(10) are in fact typical of Japanese and French syntax respectively. More generally:

(13) Japanese word order: Any projection $X^j$ including $X^0$ is head-final.

Facile comparison of Japanese with languages like French and English has led many researchers to describe the latter type as simply "head-initial" rather than "head-final." However, while French and English heads precede full phrases, they typically *follow* a range of short, non-phrasal modifiers, often termed Specifiers in earlier work; Emonds (2000: Ch. 3) provides a fuller discussion. This cross-linguistic asymmetry is already noted in the original work (Greenberg 1963) that suggests ordering parameters like (13). Thus, English and French phrasal patterns are head-initial under certain limited conditions rather than simply being the mirror image of Japanese.

Moreover, the highest projections in the cyclic domains IP, CP and DP don't conform to a head-initial pattern either: the phrases typically today called Specifiers precede the heads I', C' and D'. If we call these latter XP "closed projections" then English and French word order is not "head-initial" in closed projections either.

Finally, syntactic left-to-right order must be specified inside words, i.e. inside $X^0$. Lieber 1980, 1983 convincingly argues that within words, bound derivational suffixes are the heads whose properties determine the properties of the larger unit. Her proposal can be profitably extended to all bound morphology, provided that inflections are exempted from head-complement selection mechanisms. Then the passive suffixes are the right hand heads of the passive forms and the TENSE endings are the right hand heads of the finite verbs, as illustrated in (9)-(10). For English in particular, compounds and bound morphology are also head-final, as in Japanese.

In summary, English is not head-initial inside $X^0$, nor at the level of closed projections, nor when short modifiers precede heads. The actual extent of its head-initial patterns is more restricted:

(14) English word order: In an *open phrasal* projection $X^1$, a head precedes its phrasal sisters.

Plausibly, this extensive residue of head-final patterns in English should be assimilated to the head-final property of Japanese, making (13) redundant:

(15) Universal Right-hand Headedness. In the absence of language-particular properties (whether of syntax or morphology), heads are always rightmost among sisters.[13]

French has initial heads in a different range of structures than does English. In both phrasal and word domains, some French structures are left-headed and some are right-headed. This demonstrates that any putative syntax / morphology dichotomy is completely independent of the direction of headedness.

In French word level projections, a head can be initial under the condition that both the head and the non-head in compounds are *free morphemes.* This well-known pattern is usually over-generalised and taken to characterise all its compounds. But French compounds with at least one bound element are right-headed just like English; cf. Emonds (2000: Ch. 3) for more detail.

(16) aéroport, contre-exemple, entreprendre "undertake", kleptomane, malheur "bad luck", mi-Janvier "mid-January", monotone, motocyclette, photographie, pluridisciplinaire "multi-disciplinary", russophobie, satisfaire "satisfy", sinophile, technocrate, téléspectateur "television viewer"

Consequently, French word order differs from the English (14) as follows:

(17) French word order: In an *open projection* ($X^1$ or $X^0$), a free head precedes its free $Y^j$ sisters.

(17) predicts a second difference between French and English word order. French verbal clitics can realize phrasal complements to the left of a head, an impossible order in an English phrase. Since Romance clitics are invariably bound morphemes, clitic ordering as in (18) thus exemplifies an open projection in which a head *follows* its bound morpheme complement $Y^j$:

(18) le voir "it-see", les y mettre "them-there-put", là-dessus "there-upon", ci-joint "here-attached"

Ultimately, we would like to derive language-particular word order statements like (14) and (17) through an interaction of implicational universals of syntax with the languages' lexical or other properties, but this goal is beyond the scope of this study.

---

[13] The relation between a language-specific principle (14) and a universal (15) is the familiar Elsewhere Principle of the Sanskritist Panini formulated in generative grammar by S. Anderson and P. Kiparsky: a more specific principle (14), when applicable, overrides the more general (15).

### 3.3. The Uniform Format of Lexical Entries

Chomsky 1965 notates lexical co-occurrence of $X^0$ with a phrase by the symbol +___YP. Although I give no examples here, it can easily be appreciated that selection sometimes involves specific feature values on heads of phrases (e.g., ANIMATE, PATH, WH, etc.), and that heads themselves can be thought of as features. In any case, heads $X^0$ and their projections $X^j$ share features (by "percolation"), so co-occurrence can be better notated with +___F, where F is a syntactic feature.

Since general principles determine syntactic left-to-right order, as seen in the previous section, lexical notation should suppress the left-right ordering stipulated in the original subcategorization notation. To this end let us say that @, X, +<F> means "@ is of category X and has a phrasal complement whose head has the feature F". Thus:

(19) put, V, +<D>, +<PATH>; frighten, V, +<ANIMATE>; inquire, V, +<WH>

From these modifications, we can simplify the lexicon as follows:

(20) Lexical Interface Principle. The lexicon uses only morpheme categories, including for co-occurrence. It cannot distinguish between X and XP.

The lexicon must, however, distinguish between co-occurrence with free and bound morphemes. Lieber's 1980 word-internal subcategorization features seem appropriate for this purpose: @, X, +<___Y> means that @ of category X is a *prefix* on host category Y, while @, X, +<Y___> means that @ is a *suffix* on host category Y.[14]

## 4. Language-particular Syntactic Implications of Lexical Differences

As indicated earlier, the hypothesis (4) schematised in (6) claims that differences in category and feature content between corresponding closed class items both account for language-particular syntax and also suggest how to construct some shared LF representations. This section will show examples of how the interplay of different lexicons and universal syntax can generate characteristic aspects of language-particular syntax.

The Japanese passive morpheme *-(r)are* in (8) is a grammatical V, as its conjugation pattern shows. (An A would have the tense endings *-i* and *-katta* rather than *-ru* and *-ta*.) Therefore, the complex verbs which it heads share a universal syntactic property distinguishing V from A, namely V+*(r)are* can (but need not) assign accusative case to an object DP. Consequently, Kubo 1992 argues that the lexical category V of *-(r)are* leads to a range of Japanese transitive passives excluded in languages such as English

---

[14] With slight alterations, the lexical notation in Kubo's (8c) can conform to these notations. The superscript *k* in (8c) should be removed in line with the Lexical Interface Principle (20).

(i) Japanese passive: (r)are, V, +<V(___)>,+___$I^0$, (malefactive)

The parentheses around the context symbol in (i) serve to permit *either* morphological subcategorization (the gapped passives) or phrasal subcategorization (the gapless passives) of *-(r)are*. The same notation in (ii) indicates that *-able* is either an adjectival suffix on a V or a free adjective with a phrasal complement headed by a V.

(ii) English: able, A, +<V(___)>, POTENTIAL.

(and French), whose *adjectival* passive morphemes uniformly fail to assign structural case to objects of their host V. Her Japanese examples thus contrast with ungrammatical French word for word translations.

(21) Dative passives (the symbol t is a trace of the subject of a passive):
Taro$_i$-ga sensei-ni $t_i$ sono rekishi-no hon-o susume-rare-ta.
*Taro a été recommendé ce livre d'histoire par le professeur.
"Taro was recommended this history book by the professor"

(22) Possessive passives:
Kono sensoo$_j$-ga saisho Nihon-gun-niyotte [$_{DP}$ $t_j$ hibuta-o ] ki-rare-ta.
*Cette guerre a été tiré le feu le prémier par l'armée japonaise.
"That war has been made fire first by the Japanese army."

Taro$_j$-ga Hawai-iki-o [$_{DP}$ $t_j$ kazoku-ni ] susume-rare-ta.
*Taro a été recommendé un voyage à Hawaii à la famille.
"Taro has been recommended a trip to Hawaii by the family" (where Taro's family travels)

(23) Embedded subject passives:
Taro$_i$-ga Hanako-ni [$_{IP}$ $t_i$ sono hon-o kaita-to ] omow-are-ta.
*Taro a été cru par Hanako qu'avait lu ce livre.
"Taro has been believed by Hanako that had read that book" (where Taro read the book)

Based on examples such as (21)-(23), Kubo 1992 argues that the structural distance between the Japanese passive trace and its subject DP antecedent is not limited by case considerations, as in English. Rather, because the lexical category of the Japanese passive morpheme is a V, an NP trace in a passive is restricted only by the syntactic principle that limits maximal distance between bound anaphors (such as a passive trace) and their antecedents, namely Principle A of the Binding Theory in Chomsky (1981: Ch. 3). The differing lexical categories of the passive suffixes in Japanese (V) and English/French (A) thus lead to significant language-particular grammatical contrasts.

We may still ask, what gives rise to movement in Japanese passives if not case considerations? Kubo's answer is that some surface subject for a VP must appear in SPEC(IP), similar to the view of Åfarli 1992 on passives, widely adopted in minimalism. If this VP has as its head a grammatical verb *-(r)are*, the lexical verb heading its complement cannot skip this intermediate V and directly assign a theta role to the latter's DP subject. As a result, either *-(r)are* itself must assign an experiencer theta role to this DP (the "gapless passives"), or a DP with a theta role must move to this SPEC(IP), the "gapped passives."[15]

Kubo accounts for the difference between those *-(r)are* which assign a theta role and those which do not by analysing the former with phrasal VP sisters and the latter with

[15] According to Kubo, a main verb *-(r)are* with a VP complement assigns a malefactive theta role to its (base-generated) subject DP. The resulting "gapless" passives are the only ones that *require* an adversative interpretation. She takes it as irrelevant that many "indirect passives" (= all Japanese passives which have in common only the irrelevant property of lacking word for word English translations) can have adversative pragmatic interpretations; so can many direct passives.

V sisters. I suggest rather that both uses resemble a larger class of Japanese grammatical verbs which are independent predicates, i.e. head their own projections, but become bound suffixes on their complement at Spell Out (e.g. *-tai* "want", *-(s)aseru* "make", etc.). If so, the entry for *-(r)are* uniformly contains the feature +<V> (it takes a V-headed phrasal complement) as in (24):

(24) Revised Japanese passive: (r)are, V, +<V>, +___$I^0$, (malefactive)[16]

I propose that what distinguishes the two Japanese passives is the *level* at which *(r)are* is inserted into a derivation. The non-activity verb *-(r)are* can be inserted prior to Spell Out only if it "makes a difference" at LF, say by assigning a theta role to its subject. If inserted after Spell Out, it cannot contribute to LF directly, which forces its subject to satisfy Full Interpretation by virtue of being the landing site of some DP movement.

According to this proposal, the parenthesis notation for an inherent (as opposed to contextual) feature such as "malefactive" does not really mean that the feature is optionally present with *-(r)are.* Rather:

(25) LF Optionality. An inherent lexical feature F in parentheses does not mean "optionally present"; it means rather that F is optionally interpreted at LF.

For (24), this reinterpretation of parenthesis notation is terminological. But just below, we will see ramifications of (25) that are more than terminological.

Let's now turn to the French passive. In contrast to *-(r)are,* the French passive morpheme *-é* in (7) is an A.

(7) French passive: é, A, +<V___>, +F

By virtue of a general algorithm for morphology to be presented in section 5, this same morpheme *-é* can agree with and license a possibly empty object DP/NP position, provided it shares all the syntactic features of this DP/NP. So in order to agree fully with this object—and with no other types of phrases—the unspecified characterising feature F in the lexical entry for *-é* must be an N. This leads to the fully specified entry (7'):[17]

(7') French passive: é, A, +<V___>, N

Recall that an A cannot assign a structural case to an object DP, and so by the universal principle of the Case Filter, an object DP of a V whose form is $[_A$ V – $[_A$ *é* ] ]

---

[16] In my view, the malefactive or adversative sense in Japanese gapless passives need not be a theta role; distinguishing benefactive and malefactive theta roles seems unnecessary, especially since the two don't co-occur with the same predicate. I suggest that *-(r)are* further characterises as "negative" an otherwise neutral relation between an experiencer subject and a propositional theme, like the main verb in the archaic *Jesus suffered the little children to come unto him.* Similarly, adjunct DPs introduced by *on* in non-standard English (*My car kept stalling on me*) simply have an experiencer role, with *on* further adding an intrinsic malefactive connotation.

[17] This entry does not mean that *-é* is an N. It suggests rather that the first lexical category (here A) in an entry represents where in trees a morpheme is to be inserted; i.e. the first lexical category indicates the canonical position. I see nothing objectionable about this special status.

Since the feature N is not canonically realised in (7'), it is (correctly) not interpreted at LF.

*must* be empty. Thus, while the category A of the passive *é* in (7') prevents the French variants in (21)-(23), this same categorisation generates passive clauses whose subjects and empty objects both agree with $[_A$ *é*$]$ and hence with each other. This co-indexing is the essence of French (and English) passives.

As is well known, French and English exhibit two kinds of passives, verbal passives and adjectival passives, whose contrasting properties have been studied in generative terms since Wasow 1977. The analysis of Emonds (2000: Ch. 5) argues that all these contrasts can be explained by inserting the passive morpheme at two different derivational levels.[18] In particular, the adjectival passive corresponds to a structure in which the morpheme $[_A$ *-é*$]$ is accessible to LF and the verbal passive to one in which it is not. One correct consequence of this distinction is then that the category A, which always represents a "property" in LF, is part of the interpretation of passive adjectives but not of passive verbs.

Using LF Optionality (25), I modify (7') to make the needed distinction in insertion levels naturally fall out.

(26) Revised French passive: é, (A), +<V___>, N

According to (25), parentheses no longer mean that a category is optionally present in derivations. As its main category feature, A is present in every derivation containing *-é* from underlying structure through to Phonological Form. But parentheses now mean that A can be deleted at LF. Assuming that only empty elements are so deleted, then *-é* is inserted under A only in PF, i.e. this derives the late lexical insertion of *-é* in verbal passives argued for in Emonds (2000: Ch. 5).

The true adjectival passives, unlike verbal passives, are indeed interpreted as properties; they carry no sense of action and have no exact Japanese counterparts:

(27) a. $[_{DP}$ Les filles $]_i$ apparaissaient $[_{AP}$ atteint $[_{A,\ N,\ FEM,\ PLUR}$ es $]_i$ $[\ _{DP}$ N, FEM, PLUR $t_i\ ]\ ]$.
"The girls appeared stricken/ affected"

b. Marie et Lise se sentaient visées (the same structure).
"Mary and Lisa felt attacked/ targeted"

In (27), the subject DPs are interpreted partly by receiving the "experiencer" theta roles of the main verbs, as well as being interpreted as the objects of the dependent verbs. Since such interpretations occur with A but not with V, the different lexical categories of the Japanese and French passive morphemes explain, in conjunction with universal principles of syntax, the existence of a class of French passive structures missing in Japanese.[19] Moreover, the category A also explains why precisely interpretation as a property is the sense missing in Japanese passives.

---

[18] The analysis in question derives both types of passive via transformational movement. However, nothing changes in the argument here if adjectival passives are not so derived; the entry for the participial morpheme (26) is the guarantor of subject-object co-indexing in any case.

[19] The use of the French copular auxiliary with the passive gives rise to sentences that are ambiguous between the adjectival and verbal senses; Japanese passives lack this ambiguity.

(i) Ces filles étaient visées.
"Those girls were (in the state of) targeted" (adjectival sense)
"Those girls were being targeted" (verbal sense)

This discussion of Japanese and French suggests that any attempt to characterise each language's passive structures without centrally using the categorial specifications of the bound passive morphemes *-(r)are* and *-é* would entirely miss or misrepresent the nature of these systems. Any computational procedure that heuristically "erases" or "by-passes" the lexical specifications of bound morphology (or of other closed class items) is linguistically unreal, and inevitably fails to capture the actual combinatory properties of the individual languages.

It appears then that there are no "language-particular" grammatical rules outside of each language's lexicon of closed class items, with the possible exception of word order parameters. For another example, English has no special "particle movement" to describe the alternation in *Mary put (down) the paper (down).*[20] The only sense in which languages have their own rules of syntax is lexical: a related group of grammatical morphemes can share lexical specifications and hence constitute what is informally known as a "paradigm" (e.g., the French conjugational variants of *-é*, namely *i/u/t;* and the English directional adverbial particles).

From this perspective, it is clear that serious progress in understanding or characterising the syntax of particular languages requires analyses grounded in a theory of closed-class lexical specification. A principal result of such research is improved understanding of how and within what limits bound morphology, as well as other grammatical items, comes to mask the generalisations of universal syntax. In fact, given a proper theory of lexical formatting (section 3.3), I think a single general algorithm can recover universal bar notation representations, as discussed in section 3.1, from the widely diverse morphological systems of natural languages.

## 5. A Mapping from Morphology to Universal Syntax

### 5.1. Closed *vs.* Open Lexical Classes

In order to formulate a general morphology to syntax algorithm, the distinction between open and closed class items must first be made precise. Pre-theoretically, the distinction is obvious; languages permit at most four "open" or "lexical" categories N, A, V and P, whose members number in hundreds or thousands and may be consciously coined. In contrast, membership in the many "closed" or "functional" categories rarely if ever reaches even twenty. To represent this, I distinguish two types of features:

(28) For each bar notation category B, Universal Grammar matches a very few cognitive/ syntactic features notated F, whose combinations characterise up to at most twenty members of B.

Some examples of cognitive/ syntactic features with their canonical hosts in parentheses are: the tense and modal features (when B = I), features for quantifiers

[20] Emonds 1996 provides one non-transformational account of this alternation in terms of the lexical properties of the basic spatial particles of direction. As «late-inserted» grammatical elements, such adverbs do not appear in the terminal string at a point where the V assigns structural case to the object, and hence they do not interfere with the required adjacency.

(when B = D or possibly an additional functional head Q or NUM), the "common sense" space-time coordinates (*up, down, in, out, until, since,* etc. realised by features on P), WH (when B = C or P), perfective aspect (when B = V), and animate and count features (when B = N). In what follows, I shorten "cognitive/ syntactic" to "syntactic," keeping in mind nonetheless that syntactic features on their canonical hosts are absolutely central bearers of meaning in LF rather than "diacritics" or "purely formal".

While these syntactic features F specify broad semantic categories (e.g., ANIMATE, PAST, TEMPORAL, etc.), finer distinctions of meaning inside the four open classes are in terms of purely semantic features, notated here *f,* which play no role in syntax (Chomsky 1965: Ch. 4). Such *f* are distributed more restrictively than F:

(29) a. Only the open categories N, A, V and P may be specified for semantic features *f.*

b. A closed class item is one with no semantic feature *f,* but only syntactic features F.

Thus, all non-lexical categories are "closed" because they are never specified for purely semantic features *f.* That is, outside the lexical categories (N, A, V and P), the only elements allowed are syntactic features (F) and the small morpheme sets they characterise. Hence, such categories have few members; they moreover disallow conscious coining, which always involves semantic features *f.*

A central tenet in my approach to the lexicon concerns a certain restricted liberty of occurrence of the syntactic features F. I claim that these features can sometimes appear other than on their universally specified host, and that (only) in such cases do they become "purely syntactic" in the sense of *not* contributing to LF:

(30) Canonical Realisation. The syntactic features F that UG canonically matches to each host category B contribute to LFs only in these "canonical positions" on B. Such F appear elsewhere only by language-particular lexical stipulation and do not contribute to LF in those positions.

There is no reason to specifically exclude the four lexical categories from the scheme (29)-(30). If these categories are like the others, each has a subset of possibly say twenty elements, fully characterised by syntactic features F. Moreover, these subclasses of N, A, V and P have other properties of non-lexical classes, such as post Spell Out insertion contexts and unique syntactic behaviour (Emonds 1985: Ch. 4 and 5, 1987, 2000).[21] For example, semantically empty or light "auxiliary verbs" are simply any V not located in the canonical positions of V. We can terminologically distinguish closed subsets of lexical categories from open class items by calling them "grammatical" N, A, V and P.[22]

---

[21] These studies argue that morphemes that spell out non-canonically realised features, e.g. of agreement and case, are inserted only at PF. This hypothesis then correctly exempts right hand inflectional heads $X^0$ from interfering with head-complement selection mechanisms that involve their open class hosts.

[22] English grammatical verbs include *be, have, do, get, go, come, let, make* and probably a few others such as *put* and *say.* English grammatical nouns include *one, self, thing, stuff, other(s), people, place, time, way, reason.* The distinction between grammatical and lexical P is well known, and this scheme naturally integrates it into a general theory of grammatical categories.

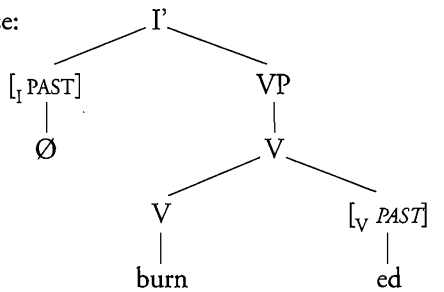
## 5.2. Alternative Realisation and Some Simple Interpretations

We have already seen a case of what I term non-canonical or "alternative realisation" of purely syntactic features. The French passive participle morpheme generated by (26) realises the grammatical category N "non-canonically" in a position of an A. The lexical entry for the English past tense (31) provides another straightforward case of Alternative Realisation; ordinarily, the canonical position for a Tense feature is on the category I, but by (31) PAST occurs on V.

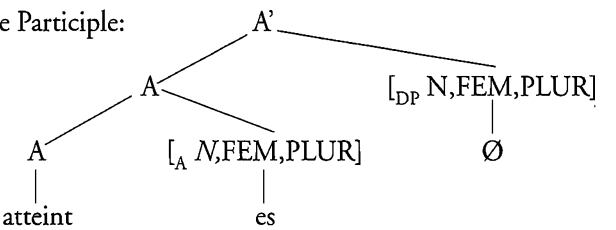
(31) Modern English: ed, V, +<V___>, PAST

These two non-canonical realisations of syntactic features are bold in the trees (32):

(32) a. English Tense:




b. French Passive Participle:




In these trees, the features PAST on V and N on A are not on their universal host categories, as determined by (30). Rather, they occur elsewhere by virtue of the language-particular lexical stipulations in (31) and (26).

The notion of canonical realisation remains contentful only if such stipulations are strictly limited. They are crucially restricted to closed class items by the following general principle:

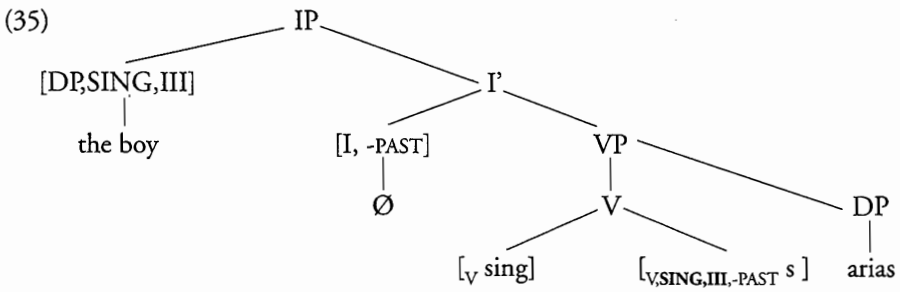
(33) Alternative Realisation ("AR"). A syntactic feature F matched in Universal Grammar with a category B can be realised in a closed class morpheme under $X^0$, provided $X^k$ is an (extended) sister of [B,F].[23]

[23] In Emonds 2000, "an extended sister of [B,F]" in (33) is generalised to "an extended sister of [B,F]$^j$." This allows for alternatively realising features of lower heads in: (i) applicative verbal morphology which realises empty Ps in PPs containing overt DP, (ii) COMP with features canonically found on I, and (iii) English "*wanna* contraction." *Wanna* is a grammatical V whose F features include an (alternatively realised) feature of *to*, but lack any features of finite *-s* or *-ed*.

(34) Extended Sisterhood. Any sister nodes W and Z are extended sisters. Moreover, if W c-commands Y and Y dominates the only lexical material under Z, then W and Y are also extended sisters.

Inspecting the trees (32) shows that the non-canonical PAST on V and N on A conform to (33) without involving the notion of "extended sister"; e.g. in (32a) PAST = F, I = B, and V = $X^0$. For an illustration of AR under extended sisterhood, consider English number agreement in e.g., *the boy sings arias*:

(35)




Since I is empty in (35), the subject DP and VP are extended sisters by (34). Hence, the canonical grammatical agreement features of DP, which are the F of (33), may then be alternatively realised in a closed class item under $V^0$, the $X^0$ of (33), since some projection of $V^0$ is an extended sister of DP. This closed class item is the agreement inflection on V.

AR provides a flexible but formally restricted characterisation of cross-linguistic lexical variation among closed class items. It essentially says that syntactic features F can appear elsewhere than where one, or we might say, LF expects. But features cannot stray too far from their universal home bases—an F can appear only on a neighbouring head, or a non-neighbouring head provided that in all intervening projections the heads are empty and the modifiers absent.

One way intervening heads can be empty is by virtue of their own features being alternatively realised. It seems that if the AR of the features F of some category B is "complete," then B can be licensed as empty. For example, AR licenses the Is in (32a) and (35) as empty, in conjunction with the following principle:

(36) Invisible Category Principle. If all marked canonical features F on B are alternatively realised by (33), except perhaps B itself, then B may be empty.

The ICP (36) is not formulated to require that a category such as B must be empty. This allows some scope for Economy of Derivation, whose function in my view is to minimise the number of language-particular insertions of free morphemes in a deriving an LF from a given underlying structure. Emonds (2000: Ch. 4) discusses the interplay between the ICP and Economy.[24]

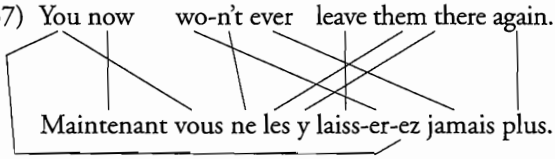
[24] Working in tandem, AR, the ICP and Economy of Derivation exclude sequences such as **did (-stress) burn, *did burned, *more tall, *more taller,* French *(le) voit lui* "(him) see him".

### 5.3. More Complicated Cases of Alternative Realisation

(32) and (35) exemplify single morphemes which alternatively realise single features. The Japanese negation illustrated in (9) may actually be another such case. Suppose that the UG host for canonical sentential negation is SPEC(VP). Then the LF for (9) should additionally contain an empty category [$_{SPEC(VP)}$ NEG]. Now, since the meaningless Japanese copula-like verb stem *-kat* does not contribute to LF, it should be inserted only at PF. Therefore in a syntactic derivation, the AP on which NEG appears (by percolation from A) is in fact an extended sister of SPEC(VP), so the NEG in (9) alternatively realises and thus licenses an empty [$_{SPEC(VP)}$ NEG]. AR can thus reduce the lexically different Japanese and French sentential negations to the same LF structures.

We do not have to look far to find multiple instances of AR within a very small syntactic space. Investigators comparing English and French cannot help but notice their overall grammatical similarity. Nonetheless, when their bound morphemes and functional categories are carefully explored, the similarity seems to vanish. For example, the left-right orderings in the following synonymous pair are essentially fixed. Lines indicate corresponding content.

(37) You now wo-n't ever leave them there again.




Such lack of correspondence might seem to jeopardise the earlier claims that synonymous structures sharing LFs involve only universal syntax and possibly shared word order parameters. But in fact, Alternative Realisation accounts for several well-known patterns of French grammar instantiated in (37), but usually discussed in terms that fail to generalise beyond Romance syntax.

For example, it is widely acknowledged that a French verb (here the stem is *laiss-*) appears in the surface position of I (like an English modal *will*) in finite clauses, by virtue of $V^0$ raising. The raising requirement can be attributed to a lexical difference in their future tense morphemes. While the English future is a free morpheme, all the French forms of Tense, which of course play a role in LF, must be in a suffixal position by Spell Out. The only way this insertion condition can be satisfied is for the V to raise in the syntactic component.

(38) English: will, I, -REALIS, +FUTURE

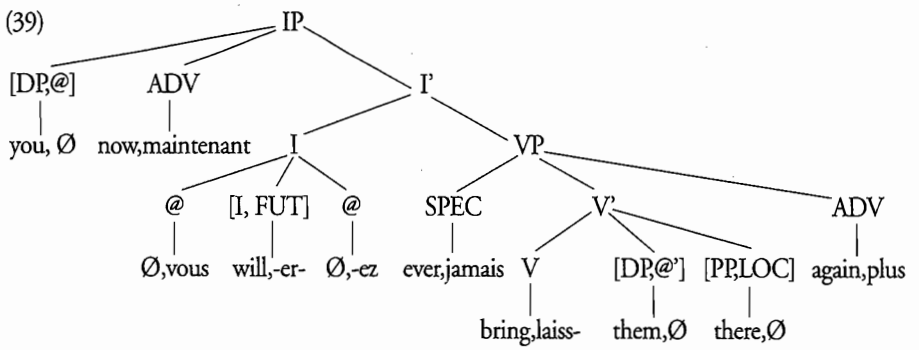
French: -er-, I, +<V___>, -REALIS, +FUTURE

More generally, the paradigm of all French finiteness morphemes taken together shares the features I, +<V___>, which insures finite verb-raising. In general, we can say that verb raising to I is a less economical option for realising V. It is a cross-linguistically available option in universal syntax, but one that occurs only if lexical specifications require it.

If we "undo" the verb movement in (37), we obtain the following underlying trees, pairing the canonically realised English and French morphemes. For brevity, let @ = II,

PLUR and @' = III, PLUR. It turns out that the same tree can essentially serve for the example in both languages.

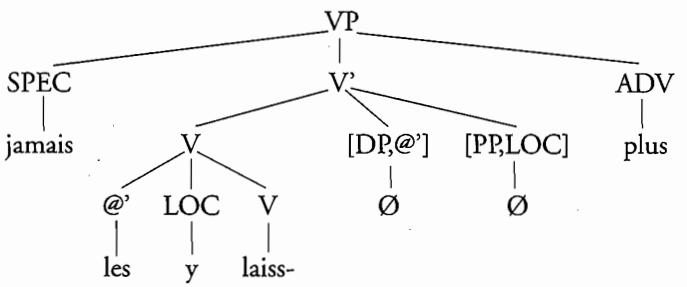
(39)




I leave aside the negative particles *n't* and *ne*. In both languages, they occur uniquely with the category I, but their ordering is different; *ne* precedes tense and modality, while *n't* follows.[25]

The only other differences between the French and English versions of (37) are the French pro-clitics *vous... les y* on the verb *laiss-*. (40) spotlights the French VP in (39), prior to V-raising to I.

(40)




It is now easy to see how the two proclitics corresponding to the two post-verbal phrases alternatively realise them. For example, in terms of (33), LOC = F, PP = B, and V = $X^0$.[26]

Apparently, AR is determined at the level of a maximal projection (say VP), and then again on a larger domain or "phase" (IP), after any intervening movement from within the smaller to the larger domain. Moreover, because alternatively realised morphemes don't contribute to LF, as stated in (30), their phonological contents are not actually present during a syntactic derivation; they are rather inserted during the

[25] Alternative Realisation may be involved here in some very local way.

[26] Emonds (1999) argues that the so-called clitic climbing in Romance causative, auxiliary, and restructuring constructions exemplifies alternative realisation of complement and adjunct phrases on the first verbs of the flat VP structures justified in the works of Rizzi 1978, Napoli 1981, Miller 1992 and Abeillé, Godard and Miller 1997.

PF sub-phases of the derivation on each successively larger domain. Thus, after AR licenses the pro-clitic positions within VP in the tree (39), V raises and left-adjoins to the contentful head [I, FUT] of IP. Then during the processing of the IP domain, further items @ which alternatively realise features of the empty subject are adjoined on the left and right of I, correctly positioning this agreement on the edges of I, in accord with Baker's (1985) Mirror Principle. This kind of bottom-up sequencing of syntactic operations agrees with a general trend of today's generative research.

In sum, the single French clause (37) exhibits at least four distinct cases of AR. The closed class morphological variations sanctioned by this principle, which include satisfying the French lexical requirement that finiteness morphemes be verbal suffixes (entailing verb-raising), account for all the discrepancies between French and English in (37). This example demonstrates how AR, while restrictively formalised, unravels the complexities of individual language deviation from canonically realised universal syntax. AR thus succeeds in confining syntactic variation to entries in closed class lexicons.

These examples reveal how certain languages can appear to downgrade or even lack grammatical categories that appear to be solid structural anchors in others. An I separate from V and obligatory overt subject DPs are hallmarks of English but not of French. Thus, even when grammatical category inventories seem similar, the language-particular syntactic realisations of corresponding morphemes can pre-theoretically appear almost chaotic. But still, cross-linguistically, morphemes of similar content typically occur "near" the same structural positions. AR is the formal device that accounts for what "near" means, and moreover explains why categories so central in some languages seem peripheral or even superficially absent in others.

This essay has claimed that all syntactic variation is ultimately reducible to two precisely formulated factors of limited scope: (i) the left-to-right ordering statements such as (14) and (17) of section 3.2, and (ii) differences in lexical items permitted by the lexical format set out in section 3.3. Canonical Realisation (30) provides basic inventories of grammatical items. The range of possible closed class items is then enlarged but still restricted by Alternative Realisation (33). The latter in particular determines exactly how far language-particular syntactic structures can diverge from the uniform LFs imposed across languages by the former.

## Bibliography

Abeillé, A., D. Godard and P. Miller, 1997, "Les causatives en français: Un cas de compétition syntaxique", *Langue Française* 115, 62-74.

Åfarli, T., 1992, *The Syntax of Norwegian Passive Constructions*, John Benjamins, Amsterdam.

Authier, J.-M. and L. Reed, 1992, "On the Syntactic Status of French Affected Datives", *The Linguistic Review* 9, 295-311.

Baker, M., 1985, "The Mirror Principle and Morphosyntactic Explanation", *LI* 16, 373-416.

Borer, H., 1984, *Parametric Syntax*, Foris Publications, Dordrecht.

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

—, 1981, *Lectures on Government and Binding*, Foris Publications, Dordrecht.

—, 1995, *The Minimalist Program*, MIT Press, Cambridge.

Emonds, J., 1985, *A Unified Theory of Syntactic Categories*, Foris Publications, Dordrecht.
—, 1986, "Parts of Speech in Generative Grammar", *Linguistic Analysis* 17, 3-42.
—, 1987, "The Invisible Category Principle", *LI* 18, 613-631.
—, 1996, "Secondary Predication, Stationary Particles, and Silent Prepositions", in A. Baba et al., eds., *Essays in Linguistics and Philology Presented to Professor Kinsuke Hasegawa*, Kenkyusha Press, Tokyo.
—, 1999, "How Clitics License Null Phrases: a Theory of the Lexical Interface", in H. van Riemsdijk, ed., *Empirical Approaches to Language Typology: Clitics in the Languages of Europe*, Mouton de Gruyter, Berlin.
—, 2000, *Lexicon and Grammar: the English Syntacticon*, Mouton de Gruyter, Berlin.
Greenberg, J., 1963, "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements", in J. Greenberg, ed., *Universals of Language*, MIT Press, Cambridge.
Kato, Y., 1985, "Negative Sentences in Japanese", *Sophia Linguistica Working Papers in Japanese* 19, Sophia University Linguistic Institute for International Communication, Tokyo.
Kayne, R., 1991, "Romance Clitics, Verb Movement, and PRO", *LI* 22, 647-686.
Keenan, E., 1975, "The Logical Diversity of Natural Languages". *Annals of the New York Academy of Sciences.*
Kubo, M., 1992, "Japanese Passives", *Hokkaido University Institute of Language and Culture Studies* 23, 231-301.
Lieber, R., 1980, *On the Organization of the Lexicon*, MIT doctoral dissertation.
—, 1983, "Argument Linking and Compounds in English", *LI* 14, 251-286.
Miller, P., 1992, *Clitics and Complements in Phrase Structure Grammar*, Garland Press, New York.
Napoli, D. J., 1981, "Semantic Interpretation vs. Lexical Governance", *Language* 57, 841-887.
Rizzi, L., 1978, "A Restructuring Rule in Italian Syntax", in S. J. Keyser, ed., *Recent Transformational Studies in European Languages*, MIT Press, Cambridge.
Stowell, T., 1981, *Origins of Phrase Structure*, MIT doctoral dissertation.
Speas, M., 1990, *Phrase Structure in Natural Language*, Kluwer Academic Press, Dordrecht.
Wasow, T., 1977, "Transformations and the Lexicon", in P. Culicover, T. Wasow and A. Akmajian, eds., *Formal Syntax*, Academic Press, New York.