Some Aspects of Word-Formation in a Polysynthetic Language
Author(s): Carolyn J. Jenkins

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via [http://linguistics.berkeley.edu/bls/](http://linguistics.berkeley.edu/bls/).

*The Annual Proceedings of the Berkeley Linguistics Society* is published online via eLanguage, the Linguistic Society of America's digital publishing platform.
SOME ASPECTS OF WORD-FORMATION IN A POLYSYNTHETIC LANGUAGE*

Carolyn J. Jenkins
The University of Chicago

Current models of word-formation (e.g. Allen (1978); Lieber (1980); Williams (1981); Kiparsky (1982, to appear)), although offering interesting insights into various aspects of the word-formation component of languages located typologically in the range from more-or-less isolating to more-or-less agglutinative, fail to account for the morphological and derivational richness of polysynthetic languages such as Greenlandic Eskimo. In this paper I will discuss the general sorts of problems raised by languages with a pervasive and totally productive system of word-derivation for any generally valid theory of word-formation. I will further argue that these problems have largely arisen from a confusion of the 'word', a surface phenomenon, with a level of linguistic analysis, an abstraction over surface regularities.

I will illustrate the basic point by focusing on a few areas where level-ordered morphology, Lieber's feature-percolation conventions, and Williams' notion of internal/external arguments make incorrect predictions about Greenlandic. That is, I will argue that these frameworks are empirically inadequate. I will not, however, conclude that they should therefore be abandoned. Rather I will argue that we need to reevaluate what Bolinger has called "our typewriter-space notions of what words are" (1971:xii) and that we may need to allow for more interaction between morphology and syntax than current, highly modularized frameworks will admit.

Level-Ordered Morphology. Kiparsky (1982, to appear) draws together several lines of investigation that have led to the general notion of level-ordered morphology. As he notes,

The basic insight of level-ordered morphology is that the derivational and inflectional processes of a language can be organized in a series of levels. Each level is associated with a set of phonological rules for which it defines the domain of application. The ordering of levels moreover defines the possible ordering of morphological processes in word-formation (1982:3).

That is, level-ordered morphology predicts a correlation between the order in which meaningful elements may occur within a word and the phonological effects those elements have on the word. In Greenlandic, however, the linear order of the components of the word is determined more by syntactic and semantic considerations than by phonological concerns.

For example, there are a number of suffixes in Greenlandic that truncate a preceding consonant. One such suffix is -lerpoq 'begin to', as illustrated in (1).¹
(1) qanik + ler poq > qanilerpoq
    approach begin to 'begins to approach'

This consonant truncation is generally considered to be morphologically conditioned; the productive pattern is for consonant-final stems to assimilate to the adjacent consonant of a suffix, as illustrated in (2).

(2) qanik + livoq > qanillivoq
    approach become more ___ 'gets closer'

Within the framework of level-ordered morphology, one would like to use the notion of the level-ordered hierarchy to obviate the otherwise unavoidable necessity of identifying truncating suffixes by a morphological feature. Indeed, this situation appears analogous to the in/-un- distinction: in-, affixed at Level One, is subject to various assimilation processes that do not apply at Level Two, the locus of un- attachment. Unfortunately for the theory, truncating suffixes may be attached not only before but also after nontruncating suffixes (illustrated in (3)), thus countering the clear prediction such an analysis makes.

(3) qanik + li(voq) + lerpoq > qanillilerpoq
    approach become more ___ begin to 'begins to get closer'

The examples in (4) indicate the role of semantics (particularly of semantic scope) in determining affix ordering irrespective of the truncating or nontruncating distinction (-ngit is truncating; -gallar is not).

(4) a. tikit + nngit + gallar + voq > tikiinnngikkallarpuq
    come not still INDIC-3s
    'he still has not come'

     b. tikit + gallar + nngit + voq > tikikkalangnilaq
    'he still hasn't come yet'/'he's a long time coming'
     (Fortescue 1980:268(6) and (7)

Another difficulty for the theory of level-ordered morphology arises from the process of lexicalization in Greenlandic. Complex forms consisting of a base and a completely productive derivational suffix often form collocations whose meaning is not transparently compositional and whose phonology may be irregular. Thus, for example, -rluk and -kulak, completely productive suffixes with straightforward phonology, occur in some lexicalized formations as instantiated in (5).

(5) a. sila + rluk > sialuk 'rain'
    weather bad (cf. sialrluk 'bad weather')

     b. angut(i) + kulak > angutikulak/angukulak
    man ugly 'he-goat' (Darden n.d.:3-4)
Although one can certainly opt to list sialuk in the lexicon, it is at least as morphologically transparent as such English forms as depth and ominous, which are held to be Level One formations. Thus we are left with three logical possibilities. First, we might remove Level One from the morphology and simply store such forms in the lexicon. In this case level-ordered morphology is essentially reduced to a distinction between derivational and inflectional morphology. Second, we could posit a basic difference between languages such as English and Greenlandic in terms of where the line between lexicon and word-formation is drawn. That is, that Greenlandic lists "Level One" formations whereas English forms them. The third alternative is to put suffixes such as -rluk in both levels and mark stems for whether they enter into formation with a particular suffix at Level One or Level Two. The problem with this approach is that most, if not all, Greenlandic suffixes are found in at least some such lexicalized or semilexicalized collocations. Thus all 350 or so productive derivational suffixes would have to be double-listed. Furthermore, as illustrated in (5), some stems follow both an unproductive and the productive pattern, with or without a shift in meaning. Thus, the general restrictiveness and predictiveness of the framework is largely undermined by this alternative analysis.

The point of the examples in (3) to (5) is not to say that level-ordered morphology cannot handle the Greenlandic data but rather that this morphological model, which is assumed to directly capture a "general property of languages" (Kiparsky 1982:11), is unrevealing with respect to one particular highly morphological language. Level-ordering simply doesn't buy us anything in Greenlandic. We cannot avoid marking various suffixes in some way for their truncating effect by locating those suffixes in a level where there is a process of truncation, nor can we uniquely assign suffixes to a particular level.

One further phenomenon in Greenlandic has direct implications for the framework of level-ordered morphology. Within this model irregular inflection may be treated as Level One affixation. Thus oxen is formed at Level One and blocks both *oxes and *oxens obligatorily. If ox is further derived—at later levels—it is irregular plural is no longer available to it, and it will be inflected regularly at Level Three. Thus, from oxishness we get oxishnesses, not *oxishnessen. It has long been observed that derived forms tend to be regular in inflection even if their base is not, and this analysis accounts for much of the data quite neatly.

Greenlandic has a form tamaq 'all of them' that has a highly irregular inflectional pattern. The usual endings are Ø or -q for the absolutive and -p for the ergative. But tamaq takes -mik and -assa as nominative and accusative endings. This situation is quite analogous to the oxen case and would presumably be treated in the same way. But if tamaq is further derived, it still requires these irregular inflections. For example, -ngajak, a fully productive suffix, may be affixed to tamaq giving the results in (6).
Furthermore, given the bracket-erasure conventions embraced by adherents of level-ordered morphology, there is no way to enable the irregular inflectional marking on tamaq to filter to the feature matrix of the derived form. It is not clear that there is any solution to this problem within level-ordered morphology other than allowing global rules. But in addition to the substantial power of global rules, they will only be effective if the internal brackets are still there for them to operate on, which presumably means retaining them in all constructions, not just those involving tamaq. Hence the attractive restrictiveness of the theory is seriously compromised if not destroyed. Moreover, it won't work to try to list these forms because they are productively derived and therefore, in principle, infinite.\(^5\)

Lieber's Percolation Conventions. Since Lieber's framework is designed to account for much the same data that level-ordered morphology is based on, both frameworks encounter many of the same difficulties when confronted with Greenlandic. The specific Greenlandic data discussed in relation to each theory represent the most natural rather than necessarily the only arguments of empirical inadequacy. These data also represent the general sorts of phenomena that will need to be taken into account in any general theory of word-formation and the lexicon.

One of the more attractive features of Lieber's framework is its ability to avoid the redundancy of stating phrase-structure rules twice—once in terms of phrase-structure rewrite rules and once in terms of strict subcategorization frames. Lieber's system "contains a single context-free rewrite rule which will generate UNLABELED binary branching tree structures" \((1980:47)\). Labeling and feature percolation conventions, which effectively ensure bracket erasure, fill out these trees as follows:

(7) FEATURE PERCOLATION CONVENTIONS:

I. All features of a stem morpheme including category features percolate to the first non-branching node dominating that morpheme.

II. All features of an affix morpheme including category features percolate to the first branching node dominating that morpheme (ibid.:49(21)).

III. If a branching node fails to obtain features by Convention II, features from the next lowest labeled node are automatically percolated up to the unlabeled branching node (ibid.: 50(25)).

Relation changing and argument adding phenomena in Greenlandic may pose a considerable problem for the percolation convention. Unfortunately, the theory, as developed thus far, is not explicit as to the analysis of such phenomena. In Lieber's
scheme, Convention III is needed to handle prefixes such as counter-, which is analyzed as having no inherent major category label but rather as being "transparent to category," taking its category label from the stem to which it is attached. That is, a tree with no major category is ill-formed, so Convention III applies. Or, in the case of Latin verbs, a verb unmarked for the features [pres] and [perf] is, in Lieber's analysis, incomplete, so Convention III applies. That is, Lieber has a notion of "feature matrices," and Convention III essentially works to ensure "that all 'empty slots' in feature matrices are filled" (ibid.: 56). In Lieber's analysis of Latin verbs, the feature [pass] is handled automatically by Convention II since the passive morpheme happens to be outermost. Thus there is no indication as to whether features indicating passivity, causativity, antipassivity are part of the feature matrix, part of the inherent content of the affix, or something else. In Greenlandic, none of these affixes is word-final, but presumably they must be visible to the syntax since they affect case marking. If these features are considered part of the feature matrix, the unlikely possibility of infinite matrices is raised. That is, the formation of antipassives, causatives, and passives productively interact with and recurse on one another to form, in principle, infinite series. (As with English sentences, semantic plausibility and human competence result in practical—though not theoretical—limits.) The examples in (8) illustrate the general point.

(8) a. Kaali Mariamiit illoqartinneqarpooq Karl(ABS) Maria-ABL house-have-cause-PASS-3s 'Karl got a house from Maria'.

    b. Kaalip Hansimut atuakkat Nuummut aggioquai Karl-ERG Hans-ALL books(ABS) Nuuk-ALL bring-ask-INDIC-3s/3pl

    'Karl asked Hans to take the books to Nuuk.'

Furthermore, such examples are extendable and further derivable. Thus one might extend (8a), for example, to Karl wanted to get a house from Maria.

Assuming infinite feature matrices to be undesirable, we are left with the option of incorporating into the morphology the syntactic apparatus of the theory (Lexical Functional Grammar) in which this morphological framework is embedded. Selkirk (1982) and Lieber (1983) both adopt this strategy to account for English compounding. On this view, argument structure and case assignment may be altered by a lexical rule associated with the derivational ending. Grimshaw and Mester (1983) explicitly analyze Greenlandic complex verbs within the framework of LFG. Without going into all the details, the analysis rests crucially on a highly debatable assumption and a serious weakening of the Functional Uniqueness Principle. The assumption is that Greenlandic is essentially a word-star language, thus enabling one to base-generate any number
of NPs in any order. But word order is more flexible than free, and one gets the proliferation of NPs illustrated in (8b) only in connection with complex, derived verbs. The Functional Uniqueness Principle is severely compromised by the stipulation that oblique grammatical relations are exempt from it. In addition to weakening the theory, this analysis completely obscures the point that, for example, the two allative arguments in (8b) do in fact bear a unique functional relation—but to different elements of a complex verb. (That is, Nuummamut is strictly subcategorized by the verb stem, and Hansimut is 'controlled' by the causative affix.) Grimshaw and Mester must weaken the Functional Uniqueness Principle for their analysis of the complex verb in Eskimo, yet this Principle, in its strong form, may be precisely right. Hence the lexical rule strategy may not be tenable.  

The major stumbling blocks for this theory, however, are certain facts of noun-incorporation. A number of verbal affixes in Greenlandic incorporate their noun object, which may be externally modified and possessed. These noun modifiers exhibit number agreement with the noun head, even though that noun "loses" its number under incorporation. The examples in (9) illustrate this construction.

(9) a. qimmeqarpoq
   dog-have-INDIC-3s
   'He has a dog/dogs'.
   (Sadock (1980:306(18))

b. angisumik qimmeqarpoq
   big-NOM-SG-INST
   'He has a big dog'.
   (ibid.:309(34))

c. angisumik qimmeqarpoq
   big-NOM-PL-INST
   'He has big dogs'.

The unmodified form in (9a) is ambiguous as to number; the modified forms in (9b) and (c) are unambiguous. Moreover, one cannot appeal to a semantic principle to account for the agreement facts because, in the case of semantically singular but formally plural nouns, there is plural agreement, thus indicating that this process is a grammatical, not a semantic one. For example, gamutit 'sled, carriage' is formally plural. In many uses it is neutral between notional singularity or plurality, just as when I say My pants are at the cleaners it is unclear whether one or more than one pair of pants is involved. But the number of sleds may of course be clarified by the use of a numeral. As shown in (10), the numeral, even if it is explicitly 'one', must have plural marking.

(10) ataatsinik /*ataatsimik qamuteqarpoq
     one-PL-INST/ one-SG-INST sled-have-INDIC-3s
     'He has one car'.
     (ibid.:309(31))
In Lieber's scheme, however, once the feature [+V] has percolated to the branching node dominating the noun-verb string, none of the nominal features of the incorporated element may be percolated up the tree to become visible to the syntax. Further, one cannot simply allow nominal features to percolate (by Convention III) in these constructions and then delete the overt marker of number and case on the incorporated noun since there are some noun-incorporating suffixes, for example, -karpoq, that do attach to inflected objects,\(^7\) as illustrated in (11).

(11) palasip illuanukarpoq (cf. illuanut 'to his house')
    priest-ERG house-3s-ALL-go-INDIC-3s\(^8\)
    'He went to the priest's house'. (ibid.:315(61))

So even if it were possible to reformulate Convention III to percolate just the right features (an ad hoc adjustment to the theory in itself), there remains the problem of deleting and expressing just the right features in the surface form.\(^9\) Thus the phenomenon of noun-incorporation appears to pose insurmountable problems for the theoretical framework.

One further point about constructions such as that in (11) is worth mentioning. Underlying most work in word-formation is the assumption that word-formation rules are structure preserving. But the phrase-structure configurations arising as a result of noun-incorporation are not otherwise occurring in the language. In particular, the ergative case occurs only (i) as subject of a formally transitive verb, and (ii) as possessor of an immediately following possessed noun. Inasmuch as the configurations of (11) occur only when an object noun has been incorporated and are simply ungrammatical otherwise, the suggestion that the word-formation component of Greenlandic is not structure-preserving is strong, if not inescapable. Moreover, it will be difficult to remove noun-incorporation from the word-formation component without also thereby controverting other favored underlying assumptions. Recent work in syntax and word-formation has been heavily focused on locating all exceptional and idiosyncratic information in the word-formation component, thus leaving only fully productive, exceptionless phenomena in the syntax. Subject to various pragmatic and syntactic constraints, it is true that incorporating suffixes may incorporate any nominal, simple or derived. But, as noted above, these suffixes do differ in terms of whether or not (and, if so, how much) inflectional material they allow on the surface—this behavior is idiosyncratic. Moreover, as illustrated in (12) (see also (8a)), affixes may incorporate derived nominals—and themselves be further derived. Hence a syntactic analysis means allowing the components to interact. Although I am arguing for this very interactivity, most theorists balk at it.

(12) illorssuaqaraluarami
    house-big-have-certainly ... but-INDIC-3s
    'He certainly has a big house, but ...'
Internal vs External Arguments. Williams (1981) argues that in addition to the argument structure of a lexical item, which he formulates in terms of thematic relations (in the sense of Gruber (1976)) whose surface form is specified by realization rules, there is "one minimal additional element of structure," that is, "the distinction between 'internal' and 'external' arguments" (1981:82). The internal/external distinction is intended to replace, not supplement, grammatical relations. "The notion external argument corresponds in some respects to subject," but is conceived of as "simply a distinguished argument, not a syntactic position, a case, or something else" (ibid.:83).

The real substantive claim of Williams' article is in terms of the restriction he is able to place on possible morphological rules. Specifically, Williams claims that "a morphological rule can affect only the external argument of its input, and that it can affect this external argument in only one of two ways: it can make one of the internal arguments into the external argument, or it can add a new external argument" (ibid.:90). This restriction (coupled with, for example, the requirement that the Actor, if there is one, be external) constrains analysis of many linguistic phenomena. Specifically, verbal passives must be analyzed as morphological and antipassives as syntactic. As shown in (13a) and (b), the antipassive changes case marking but not argument structure—Hansi is both Actor and external argument in both constructions.

(13) a. Hansip illu sanavaa
     Hans-ERG house(ABS) build-INDIC-3s/3s
     'Hans built the house'.

   b. Hansi illumik sanavoq
     Hans(ABS) house-INST build-INDIC-3s
     'Hans built a house'.

A question, of course, naturally arises: is it correct to analyze the passive as morphological and the antipassive as syntactic? The passive is morphologically more regular in that the passive suffix -negar is the same for all verbs, whereas Klein-schmidt (1851) notes several suffixes that do "not change the meaning of the stem in other ways than to relieve it of the transitive suffixes" (p. 73 of Anderson's translation). On the other hand, whereas passivization always requires an overt derivational affix, antipassivization for many verbs is signalled simply by the intransitive inflection on the surface form. Both processes are fully productive: every formal transitive verb has both a passive and an antipassive.10 Thus both processes are equally syntactic-like in terms of productivity.

Since this discussion is not conclusive, I will simply assume for the moment that passive is lexical and antipassive syntactic and that, as Williams claims, this analysis is the only possible one. But if Williams is right then the syntax and the
word-formation component must interact because passive and anti-passive clearly do interact. The examples in (14) are taken from Inuktitut (a Central Artic dialect of Eskimo), but similar examples from the Greenlandic dialect can be constructed.

(14) a. nutaraq arnarmit titirautimik nanisirqujauvuq
    child(ABS) woman-ABL pencil-INST find-ANTP-tell-PASS-
    INDIC-3s
    'The child is/was told by a/the woman to find a/the
    pencil'.
    (Johnson 1980:25(53))

b. Piita Maa limik Jaanimit kuniktuarqujivuq
    Peter(ABS) Molly-INST Johnny-ABL kiss-PASS-tell-ANTP-
    INDIC-3s
    'Peter tells/told Molly to be kissed by Johnny'.
    (ibid.:25(57))

These examples do not, of course, exhaust the derivational wealth of Eskimo dialects. They do, however, serve to illustrate the basic point of the interactivity of these processes: the lexical passive and the syntactic antipassive may apply to one another (as well as apply to their own output).

How this interaction is to be formulated remains a question, but that it must occur seems to be required by the data. Looking just at the verbs in (14a) and (b) two possibilities suggest themselves immediately. First, one could conceive of Greenlandic word-formation (and, presumably English sentence-formation) as a grand spiral of (perhaps level-ordered) successive recursions through a cycle of syntax-morphology-phonology coupled by semantic interpretation. Second, one might try to build up these complex Greenlandic forms syntactically in several pieces that are then glued together through the word-formation component. For example, in the formation of the verb in (14b), one might start with 'kiss' and 'tell' in the syntax; 'kiss' would immediately be sent to the word-formation component for passivization whereas 'tell' would remain in the syntax to be antipassivized. The two forms would ultimately join together morphologically and be sent forward for inflection. The problem with the first possibility is that it is difficult to imagine how such a spiral might be constrained. The problem with the second is that the syntax still needs to know that 'kiss' has been passivized (so jaani may be assigned ablative case) so it is not certain that the interaction can be limited to the effect of morphologically joining syntactically constructed elements.

Both these possibilities involve a certain expansion of the interactivity of the linguistic components. And both pose the problem of then restricting that interactivity on principled grounds. I am suggesting that it might be useful to explore a more abstract notion of the 'units' of our inquiry. In particular, it seems inappropriate to equate the surface word with a single linguistically relevant unit that is necessarily the same at
all levels of analysis. Presumably there is agreement among the components in most cases. For example, English woman and Greenlandic arnaq 'woman' constitute a unit for the semantics, syntax, morphology, and phonology. But in Greenlandic at least there does seem to be a distinction among 'contentful' suffixes (such as 'tell', 'want', 'go to'), category-changing suffixes (various nominalizers and verbalizers not discussed here), relation-changing suffixes (such as the passive and antipassive affixes), and inflectional endings. The contentful affixes appear to be semantic, and in some cases also syntactic, units. The relation-changing affixes seem to be units of the syntax. On this view, the surface word is the result of the usually confluent but sometimes competing demands of the several components of the grammar.

Footnotes

* Sincere thanks to Jerry Sadock for much valuable discussion and for prodigious assistance with the Greenlandic examples. Thanks also to Bill Darden and the Linguistic Circle of Chicago for comments and encouragement. Alas, any elements of fiction or fantasy that have crept into this manuscript are my responsibility.

1 References for examples are given within the example itself; the source example numbers are in parentheses after the page number. Where there is no citation, the form has been provided verbally by Jerry Sadock. Some examples have been revised slightly to be consistent with the format and abbreviations employed in the rest of the text.

2 Rischel (1974) illustrates that "attempts to predict the status of suffixes as truncating or non-truncating from phonological or semantic properties are likely to encounter serious difficulties" and therefore "that this property must be indicated as a morphological feature" (1974:197-8). Thus, for example, there are the phonological near minimal pairs -li 'but' (nontruncating) vs. -lik 'provided with __' (truncating) and -niq 'the act/state of __' (nontruncating) vs. -ni 4th person (truncating). Two suffixes taken from the same semantic domain but differing in this respect are -taaq 'new' (nontruncating) vs. -tuqaq 'old' (truncating).

3 The terminology here is somewhat misleading. Level-ordered morphology is not committed to the proposition that the output of Level One and Level Two is derived in any dynamic, on-line sense. Indeed, the theory may be viewed as a means of formulating lexical redundancy rules as a component of the grammar (the wordformation component) and of providing these redundancy rules with a theoretical framework that increases the amount of work they are able to do in the grammar. Thus the distinction between these two analytical options is actually greater than that indicated in the text since option two implies that English relates lexical items
through the principles of level-ordering (in the word-formation component) whereas Greenlandic relates them through redundancy rules (in the lexicon).

Although there may be ways of weakening the theory to describe the Greenlandic data (Paul Kiparsky suggested one way at the BLS Conference), it seems more appropriate to relax the claim for universality of the model than to deprive the framework of its essential restrictiveness.

Given the meaning of tamaq the possible forms based on it may not in fact be infinite but only indeterminately large. It remains the case, however, that the set is productively formed with compositional semantics and regular phonology and that the language could enlarge that set if, for example, it acquired a new suffix with appropriate semantics or a new concept that it could name through existing resources.

Joan Bresnan noted at the Conference that it is possible within LFG to build up complex functional structures for complex verbs. I have not, however, had an opportunity to evaluate such an analysis.

The attachment of derivational affixes to inflected forms is explicitly ruled out by level-ordered morphology as well as by Lieber.

Analyses of this construction differ. Some (e.g., Sadock) analyze the affix as -karpoq and the -mu- and -nu- preceding it as the truncated forms of the allative ending (-mut sg and -nut pl). Others (e.g., Fortescue) argue that -mukarpoq and -nukarpoq are allomorphs of a single incorporating affix. A problem with the latter analysis is in terms of capturing the distributional pattern of -mukarpoq and -nukarpoq. At any rate, the 3s marking on illu (to agree with the singular possessor) is clearly inflectional.

This area is one in which level-ordered morphology may provide an option for analysis. That is, one could posit post-inflectional levels to deal with suffixes such as -karpoq. But such an approach may result in a potentially infinite number of levels since there appears to be a "clitic cline" in Greenlandic such that affixes grade along a continuum from clearly bound derivational suffix to full word status (Sadock 1983).

The one surface exception is that modal notions such as 'want' can be neither passivized nor antipassivized, although, of course, its verb complement can be.

Bibliography


