Recursive Compounds and Word-Stress Location

Hisao Tokizaki
Sapporo University
toki@sapporo-u.ac.jp

1. Introduction

In this paper, I argue that the complexity of compounding is correlated to word-stress location in languages. I show that recursive compounds are not found in languages with right-hand stress such as right-edge and penultimate stress. I argue that this is because the movement of complement into specifier position would make left-branching compounds with left-hand stress, which conflicts the right-hand word-stress in the languages.

In section 2, I consider the typology of recursive compounds and propose the distinction between single-head and multi-head recursive compounds. It is pointed out that the latter has left-branching structure and are found only in left-hand stress languages. In section 3, I argue that the compound-movement functions as compounding in that it changes right-branching structure into left-branching, which has strong juncture between elements. In section 4, I propose a phonological constraint on complement-movement to the effect that word-stress location matches the main stress location of compounds derived from compound-movement. We can explain the recursivity of compounds in languages in terms of the universal constraint and the language-specific word-stress location.

2. The typology of recursive compounds

Haider (2001) argues that recursive compounds (complex compounds in his term) are possible only if their structure is head-final. He discusses the difference of compounding between Germanic such as English and Romance languages such as French, as shown in (1) (the bracketing is his own).

(1) a. [baby [cat [fish]]]
b. [[[*poisson*] *chat*] (**bébé*)]
fish cat baby

He argues that the differences stem from the basic branching condition (BBC) to the effect that each branching node on a projection line is to the right of its sister node.

(2) BBC: Each branching node on a projection line is to the *right* of its sister node.

The effect of BBC can be shown schematically as in (3), where A, B and C show words and H the head of the whole constituent.

(3) a. [A [B [C H]]]
b. * [[[H A] B] C]

Haider claims that (1a) observe BBC because they have right-branching structure (3a). The three-word compound in (1b) has left-branching structure (3b), which is ruled out by BBC.

Haider's analysis gives us an interesting point of view of the recursivity in compounds. However, it also has some problems and raises some questions. First, the examples he discusses as recursive compounds can be considered to be phrases whose first word functions as an attributive adjective. For example, the structure of (1a) can be (4).

(4) [$_{NP}$ [$_{A}$ *baby*] [$_{N}$ *cat fish*]]
'small catfish'

In fact, some dictionaries list *baby* as an adjective as well as a noun (e.g. *Oxford Advanced Learner's Dictionary* and *Collins Cobuild*). In section 3, I argue that right-branching structure such as (4) is a phrase rather than a compound.

Second, head-initial recursive compounds can be found in Italian (Dressler 1988).

(5) a. *sala personale viaggiante*
room staff traveling
'traveling staff room,'
b. *sala dirigente capo*
'chief-executive room'

Scalise (1992: 196) notes the possibility that these examples can be abbreviated phrases or phrasal compounds. We will discuss the problem of compound/phrase distinction in section 4.[1]

Third, there are a large number of left-branching compounds in certain types of languages.

(6) a. [[[*waste*] *disposal*] *plan*]
b. [[[*towel*] *rack*] *designer*]

Mukai (2008: 191, 193) argues that right-branching compounding is more restricted than left-branching compounding for some reason. She uses the mark # to show that the example do not exist although it is grammatical.

---

[1] Baroni et al. (2007) also argue that deverbal compounds in Italian are phrase-like in that they are easily extended by recursion, even when they are modified:

(i) a. *Raccolta privata rifiuti tossici*
collection private waste toxic
b. *delibera comunale raccolta privata rifiuti tossici*
regulation municipality collection private waste toxic

(7) a. # [*kodomo* [*hon kurabu*]]
child book club
'book club for children'
b. # [child [book club]]

(8) a. [[*kokka kooan*] *iinkai*]
nation safety committee
'the National Public Safety Commission'
b. [[theatre ticket] shop]

She observes that it is hard to construct right-branching compounds with more than three roots.

Haider's claim might seem to contradict Mukai's observation. However, I argue that both of their observation can be right because they discuss different types of 'recursive compounds.' Haider's example (1) contains only one head *fish*, which is merged with a modifying noun *cat* to make a compound *cat fish*, which in turn is merged with a modifying noun (or adjective) *baby* to make a 'recursive compound' *baby cat fish*. Mukai's examples in (7) and (8) as well as (5) and (6) contain two heads. To take (6a) for example, the first word *waste* is merged with the deverbal head *disposal* to make the first compound *waste disposal*, which is merged with the second head *plan*.

The distinction between single-head recursive compounds and multi-head recursive compounds reveals the typology and universals of the recursivie compounds. I propose the generalizations in (9) and (10).

(9) a. Single-head 'recursive compounds' are right-branching.
b. Multiple-head recursive compounds are (consistently) left-branching.

(10) a. Single-head 'recursive compounds' are found in any languages.
b. Multiple-head recursive compounds are found in certain languages.

The fact (9a) might be explained by Hider's BBC, although it simply stipulates the possible direction of branching. The fact (9b) corresponds to Mukai's observation of the contrast between right-branching (7) and left-branching (8) compounds with multiple heads. The typological facts in (10a) and (10b) have never been discussed in literature as far as I know. Below I will show that the contrast between (10a) and (10b) holds as an universal in languages. Then, I will give a principled answer to the question why compounds show the facts in (9) and (10).

Let us now consider the typological difference between single-head 'recursive compounds' (10a) and multiple-head recursive compounds (10b). For example, Scalise (1992: 196) points out that Italian compounds are not recursive while in English, Dutch and German (i.e. in Germanic languages in general) compounding is normally recursive (cf. Selkirk 1982:15, Booij 2009b: 205).

(11) a. [[towel rack] designer]
b. [[[towel rack] designer] training]

(12) [[[*ziekte verzuim*] *bestrijdings*] *prgramma*]
illness absence fight programme
'programme for reducing absence due to illness'

In Greek, only right-branching compounds can be recursive (cf. Ralli 2009: 457).

(13) [*meγal-* [*kapn-* *emboros*]]
big tobacco merchant
'big tobacco merchant'

Mukai (2008) shows a table of languages which allow recursive compounds and which do not. Bauer (2009: 350) also notes that there appear to be languages which do not permit recursion, including Slovak, and others such as Fongbe and Ngiti (Sudanic) where recursion is extremely limited. However, these studies do not clearly show the recursivity of multiple-head compounds because they do not distinguish them from single-head compounds. We need to investigate which languages allow multiple-head recursive compounds.

Assuming Kayne's (1994) universal base structure with Spec-Head-Complement order, I argue that left-branching compounds are real recursive compounds, in which the complement iteratively moves to the specifier position of the head, as shown in (14) (For the compound nature of left-branching structure, see Wagner 2005 and Booij 2009).

(14) [plan [disposal [waste]]] → [plan [[*waste*] disposal]] → [[[*waste*] *disposal*] plan]

In (14) the lowest complement *waste* moves to the Spec of *disposal* to make a compound *waste disposal*, which moves to the Spec of *plan* to make a recursive compound *waste disposal plan*.

3. Left-branching structure as compounds

In this section, I argue that left-branching compounds are real recursive compounds while right-branching compounds are phrases. The distinction is due to the asymmetry of juncture between left-branching and right-branching structure (Tokizaki 2008, Tokizaki and Kuwana 2010).[2]

Tokizaki (2008) argues that juncture between constituents is shorter in left-branching structure than in right-branching structure. Let us consider the structures in (15).

(15) a. [[X Y] Z]
b. [X [Y Z]]

[2] This section contains the arguments which have appeared in Tokizaki (2008) and Tokizaki and Kuwana (2010). These are shown here because they are the base for the analysis of recursive compounds to be presented below.

The two structures in (15a) and (15b) are symmetrical in constituency, but they differ in phonological realization. The evidence is based on phonological changes in compounds. Japanese Sequential Voicing (*Rendaku*) and Korean *n*-Insertion occur between constituents in left-branching structure but not in right-branching structure, as shown in (16) and (17) (cf. Otsu 1980, Han 1994).

(16) a. [[*nise tanuki*] *shiru*] → *nise danuki jiru*
mock badger soup
'mock-badger soup'

b. [*nise* [*tanuki shiru*]] → *nise tanuki jiru (*danuki)*
mock badger soup
'mock badger-soup'

(17) a. [[*on chən*] *yok*] → *on chən nyok*
hot spring bathe
'bathing in a hot spring'

b. [*kyəŋ* [*yaŋ sik*]] → *kyəŋ yaŋ sik (*nyaŋ)*
light Western food
'a light Western meal'

Left-branching structure in (16a) and (17a) allows phonological rules to apply between constituents, and right-branching structure in (16b) and (17b) does not. This fact can be explained by the assumption that left-branching structure has shorter juncture between its constituents than right-branching structure. The long juncture between constituents blocks phonological rules in (16b) and (17b).

The role of junctural asymmetry in compounds is further supported by data in Dutch. Krott et al. (2004) show that in Dutch, the occurrence of interfixes including *-s-* in tri-constituent compounds matches the major constituent boundary better in right-branching compounds than in left-branching compounds. They counted the occurrences of tri-constituent compounds in the Dutch section of the CELEX lexical database, which is based on a corpus of approximately 42 million words. In (18) and (19), the number of compounds with *-s-* and all interfixes are shown in parentheses after the examples.

(18) a. Interfixes at the constituent break in right-branching compounds (*-s-* 38; all 60)
[*arbeid-s-*[*vraag stuk*]]
employment+question-issue

b. Interfixes within the inner compound in right-branching compounds (*-s-* 3; all 11)
[*hoofd* [*verkeer-s-weg*]]
main+traffic-road

(19) a. Interfixes at the constituent break in left-branching compounds (*-s-* 25; all 39)
[[*grond wet*]-*s*-*aartikel*]
ground-law+article, constitution
b. Interfixes within the inner compound in left-branching compounds (*-s-* 13; all 50)
[[*scheep*-*s*-*bouw*] *maatschappij*]
ship-building+company

The ratio of the unmarked interfix position (18a) and (19a) to the marked interfix position (18b) and (19b) is higher in right-branching (18) (*-s-* 38÷3=12.7; all 60÷11=5.5) than in left-branching (19) (*-s-* 25÷13=1.9; all 39÷50=0.8). Comparing the ratios between unmarked and marked interfixes of right-branching 12.7 (*-s-*) and 5.5 (all) with those of left-branching 1.9 (*-s-*) and 0.8 (all), we conclude that interfixes in the unmarked position are more likely to occur in right-branching compounds than in left-branching ones. That is, interfixes occur at the constituent break more often in right-branching compounds than in left-branching compounds. In other words, interfixes are more likely to occur within the inner compound in left-branching compounds than in right-branching compounds. This result is expected if we assume that the juncture between constituents in right-branching structure is long enough for interfixes to intervene there. In left-branching structure (19), the juncture between the second word and the third is about as short as the juncture between the first word and the second. Thus, marked interfixes (19b) can occur more frequently in left-branching structure (19b) than in right-branching structure (18b). This fact supports our junctural asymmetry hypothesis.

Moreover, junctural asymmetry can also be seen in morphology. Hyman (2008: 323) argues that suffixes tend to be more tightly bound to their stem than prefixes. This observation also supports the asymmetry in juncture hypothesis because [$_{\text{Word}}$ prefix [$_{\text{Stem}}$ ...]] is right-branching while [$_{\text{Word}}$ [$_{\text{Stem}}$ ...] suffix] is left-branching. Thus, cross-linguistic facts show that juncture between constituents is longer in right-branching structure than in left-branching structure.

Furthermore, there is also syntactic evidence for the asymmetry between the left-branching structure and the right-branching structure. Wagner (2005) shows that there is a phrasing asymmetry between OV and VO orders: OV is pronounced as a prosodic phrase while VO is pronounced as two prosodic phrases. In (20), parentheses show prosodic phrases.

(20) a. (*Sie hát*) (*einen Tángo getanzt*)
she has a-Acc tango danced
'She has danced a tango.'
b. (*Sie tánzte*) (*einen Tángo*)
she danced a-Acc tango
'She danced a tango.'

The OV in (20a) [[*einen Tángo*] *getanzt*] is left-branching and is included in a prosodic phrase. The VO in (20b) [*tánzte* [*einen Tángo*]] is right-branching and is divided into different prosodic phrases (cf. Wagner 2010 for branching direction in coordinate structure). It is also

to be noted that Booij (2009a) also argues that some OV sequences in Dutch behaves like compounds.

Finally, the typology of adverbial subordinators also supports the junctural asymmetry hypothesis. Investigating 611 languages in the world, Dryer (2005) points out that all clear instances of affixal adverbial subordinators (Sb) are suffixes on the verb, with no clear instances of prefixes on the verb.

(21) a. $[_{CP}$ Sb $[_{IP}$ ...]] (367 languages)
b. *$[_{CP}$ Sb-$[_{IP}$ ...]] (0 languages)
(22) a. $[_{CP}$ $[_{IP}$ ...] Sb] (90 languages)
b. $[_{CP}$ $[_{IP}$ ...]-Sb] (59 languages)

The fact that (21b) does not exist shows that clause-initial adverbial subordinators must be separated from IP as in (21a). This is because CP is right-branching in (21) with its immediate constituents, adverbial subordinator and IP, separated from each other by the long juncture between them. It is impossible to attach an adverbial subordinator to the following IP as a prefix. Clause-final adverbial subordinators can be attached to the preceding IP as a suffix as shown in (22b) because they merge with the IP on its left to make a left-branching structure. This is possible because the juncture between constituents in left-branching structure is short enough for adverbial subordinators to attach to the preceding clause. Thus, the data in (21) and (22) support the junctural asymmetry hypothesis.

All these arguments support the idea of left/right-branching asymmetry. In the next section, I argue that the shorter juncture in left-branching structure works as compounding words, which is constrained by the unmarked stress location in the language.

4. The stress location in recursive compounds

4.1. Word-stress location and head-complement orders

Next, I argue that complement-movement applies only if the resulting compound has the unmarked stress location of a simplex word in the language. Goedemans and van der Hulst (2005) classify the stress location types into two categories, fixed stress and weight-sensitive stress. Fixed stress languages put stress on the fixed position of a word such as ultimate, penultimate, antepenultimate and initial. Languages with weight-sensitive stress put stress on a certain range of positions depending on the syllable weight (heavy/light). These languages are divided into some types, e.g. Right-edge (ultimate and penultimate), Right-oriented (ultimate, penultimate and antepenultimate).

Tokizaki and Kuwana (2010) argue that stress location correlates the order of head and complement in languages. As stress location goes from right to left, the XP with complement-head order gets larger. For example, penultimate stress languages such as Bantu and Right-edge stress languages such as Romance have complement-head order only in simplex words (Stem-Affix). Right-oriented languages such as Germanic have complement-head order in words (Stem-Affix), compounds (complement word-head word), noun phrases (Genitive-N). Initial stress languages such as Uralic have complement-head

order in PP (DP-Postposition) and VP (OV) as well as words, compounds and noun phrases.

The correlation between stress location and head-complement orders is explained by a stress constraint on XPs with complement-head order derived by complement movement into specifier position of a head. XPs with complement-head order are like compounds with strong juncture between elements inside them, and should have the same stress location as a simplex word in the language.

4.2. Complement-head order in recursive compounds

Now let us consider the structure and recursivity of compounds.
Complement movement would make a left-branching compound with the stress on the leftmost constituent (stress shown with underscore, an accent mark (´) or bold face below).

(23) $[_{HP}\ [_{H}\ \ldots]\ [_{Complement}\ \ldots]]$ → $[_{HP}\ [_{Complement}\ \underline{\ldots}]\ [_{H}\ \ldots]]$

The derived constituent is like a compound word as we have discussed above. Then its stress location should match the word-stress location in the language. This stress constraint is satisfied in Germanic and Uralic languages, which have Right-oriented and initial stress.

(24) a. master station → státionmaster
b. disposal waste → wáste disposal

(25) a. *főnök állomá* → *állomásfőnök*
boss station station-master (Hungarian)
b. *päällikkö asema* → *asemapäällikkö* (Finnish)
master station station-master

The derived constituents have left-hand stress, which is allowed in these languages. This is not the case in Romance and Bantu languages, which have Right-edge and penultimate stress.

(26) a. *capo-stazióne* → **stazi**ó**necapo* (Italian)
head station
'stationmaster'
b. dereva teksi → *t**e**ksi dereva (Swahili)
driver taxi
'taxi driver'

The derived forms would violate the stress constraint in these languages. They would have stress on the complement according to Cinque's (1993) stress rule, which assigns stress on the most deeply embedded element. However, this stress location does not conform to the right-edge stress. Thus, Romance and Bantu languages allow only quasi-compounds without complement-movement shown in the left column in (26).

These quasi-compounds are not compressed enough because they have right-branching structure, which have weak juncture between elements. They are not compressed enough to be a part of another compound as shown in (27a=1b) and (27b).

(27) a. * [*poisson* [*chat* [*bébé*]]]
fish cat baby
b. * [*gatto* [*capo stazióne*]]
cat master station
'a cat who works as the stationmaster'

Thus, recursive compounds are not found in languages with right-edge stress and righthand fixed-stress such as penultimate stress (e.g. Polish: Szymanek 2009: 477)) and antepenultimate stress (e.g. Greek: Mukai 2008). On the other hand, complement-movement can make recursive compounds in languages with initial stress (Finnish and Hungalian: Mukai 2008) and right-oriented stress, which allow left-hand stress such as antepenultimate stress. For example, recursive compounds such as *wáste disposal plan* and *státionmaster cat* have stress on the first constituent, which can be allowed in such languages.

If a language does not have word-stress, nothing prevents the complement-movement from making recursive compounds. Thus, we can explain the fact that tone languages such as Chinese, pitch accent languages such as Japanese and Kyungsan Korean and no accent languages such as Seoul Korean have recursive compounds.

(28) a. *gomi shori keikaku* (Japanese)
waste disposal plan
b. *fèiwù chǔzhì jìhuà* (Chinese)
waste disposal plan
c. *sseulegi cheoli gyehoeg* (Korean)
waste disposal plan

These languages have no stress and allow complement-movement to a spec of a head to make a left-branching compound.

Note that the analysis presented here predicts that languages with word stress only allow compounds with the unmarked word-stress location. If a language has no stress, the analysis predicts that the stress constraint does not prevent complement-movement from making compounds. However, there is a possibility that compounding is also ruled out by some other constraints, which may be phonological or morphosyntactic. In that sense, we need to investigate compounding in tone languages, which does not use stress.

5. Conclusion

I have argued that word-stress location determines the recursivity of compounds in a language. It is not always clear whether a language has recursive compounds or not (cf. Namiki 2002; Mukai 2008 for a list of languages with recursive compounds). We need to investigate more languages, especially tone languages without word-stress. However, this theoretical approach to recursive compounds gives us a new working hypothesis on the morphosyntax-phonology interface and the typology of compounding in languages.

References

Baroni, Marco, E miliano Guevara and Roberto Zamparelli. 2007. Italian deverbal compounds: Words, phrases or either? Paper presented at 33esimo Incontro di Grammatica Generativa, University of Bologna.

Bauer, Laurie. 2009. Typology of compounds. In Rochelle Lieber and Pavol Štekauer (eds.), 343-356.

Booij, Geert. 2009a. A constructional analysis of quasi-incorporation in Dutch. *Gengo Kenkyu* 135, 5-27.

Booij, Geert. 2009b. Compounding and construction morphology. In Rochelle Lieber and Pavol Štekauer (eds.) *The Oxford handbook of compounding.* Oxford: Oxford University Press, 201-216.

Dryer, Matthew. 2005. Order of adverbial subordinator and clause. In Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) *The world atlas of language structures*, Oxford: Oxford University Press, .

Goedemans, Rob and Harry van der Hulst. 2005. Weight-sensitive stress. In Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) *The world atlas of language structures*, Oxford: Oxford University Press, 66-69.

Haider, Hubert. 2001. Why are there no complex head-initial compounds? In Ch. Schaner-Wolles, J. Rennison and F. Neubart (eds.) *Naturally!: Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*. Torino: Rosenberg & Sellier, 165-174.

Hyman, Larry M. 2008. Directional asymmetry in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46, 309-350.

Kayne, Richard S. 1994. *Antisymmetry in syntax*. MIT Press.

Krott, Andrea, Gary Libben, Gonia Jarema, Wolfgang Dressler, Robert Schreuder, and Harald Baaven. 2004. Probability in the grammar of German and Dutch: Interfixation in triconsonstituent compounds. *Language and Speech* 47, 83-106.

Lieber, Rochelle and Pavol Štekauer (eds.) 2009. *The Oxford handbook of compounding*. Oxford University Press.

Mukai, Makiko. 2008. Recursive compounds. *Word Structure* 1, 178-198.

Namiki, Takayasu. 2002. Go-o tsukuru shikumi: Keitairon I (Word-formation system: Morphology I). In Yukio Otsu et al. (eds.) *Gengo-kenkyu-nyumon* (An introduction to the study of language). Tokyo: Kenkyusha, 76-88.

Ralli, Angela. 2009. IE, Hellenic: Modern Greek. In Rochelle Lieber and Pavol Štekauer (eds.) *The Oxford handbook of compounding.* Oxford: Oxford University Press, 453-463.

Scalise, Sergio. 1992. Compounding in Italian. *Rivista di Linguistica* 4, 175-199.

Selkirk, Elisabeth. 1982. *The syntax of words.* Cambridge, MA: MIT Press.

Szymanek, Bogdan. 2009. IE, Slavonic: Polish. In Rochelle Lieber and Pavol Štekauer (eds.), 464-477.

Tokizaki, Hisao. 2008. Symmetry and asymmetry in the syntax-phonology interface. *Phonological Studies* 11, 123-130.

Tokizaki, Hisao and Yasutomo Kuwana. 2010. A stress-based theory of disharmonic word orders. To appear in Theresa Biberauer and Michelle Sheehan (eds.) *Theoretical approaches to disharmonic word orders*. Oxford: Oxford University Press.

Wagner, Michael. 2005. Asymmetries in prosodic domain formation. *MIT Working Papers in Linguistics* 49, 329-367.

Wagner, Michael. 2010. Prosody and recursion in coordinate structures and beyond. *Natural Language and Linguistic Theory* 28, 183-237.