Productivity of recursive compounds

Makiko Mukai, University of Kochi, Japan

Recursion at word-level is productive in many languages across the world, just as it is at phrase-level (Roeper et al. 2002; Bisetto 2010). The standard assumption is that left-branching recursive compounds (e.g. [[student film] society]) are more productive than right-branching recursive compounds (e.g. [student [film society]]) (Mukai 2008, 2017; Tokizaki 2011; Pöll 2015). However, this assumption has hardly ever been tested empirically in more detail. Using Corpus of Contemporary American English and National Web Japanese Corpus for Japanese and native speaker's judgements on the semantic interpretation of the data, we found that the prediction is borne out. In other words, recursive compounds are preferably interpreted as left-branching. After presenting representative data from each language I will propose that left-branching recursive compounds [[A B] C] are easier to parse, since a constituent can be formed earlier than in [A [BC]] structures (Pöll 2015).

Keywords: *recursion*, *left-branching recursive compounds*, *right-branching recursive compounds*, *parse*, *corpus study*

1. Introduction

The standard assumption is that LEFT-BRANCHING RECURSIVE COMPOUNDS [[AB] C] are easier to parse, since a constituent can be formed earlier than in RIGHT-BRANCHING RECURSIVE COMPOUNDS [A [BC]] (Mukai 2008, 2014; Pöll 2015). However, this has hardly been tested empirically. In this paper, the main aim is to test the hypothesis that right-branching recursive compounds are more restricted in frequency in English and Japanese, with the assumption that the higher the token frequency of word the higher the productivity of the word is (Baayen & Lieber 1997). At the end of this paper, some potential reasons behind the restriction of right-branching recursive compounds (Pöll 2015) will be discussed.

RECURSION is said to be a fundamental property of human language that potentially differentiates it from both other human cognitive domains and known communication systems in animals (Hauser et al. 2002; Corballis 2011). So, to reveal the characteristics of recursion at word-formation can reveal some aspects of human language. Recursion has been questioned in one of the Amazonian languages, Pirahã (Everett 2005). Everett (2005: 5) gives an example of embedded clauses in English where in Pirahã it is expressed without embedding. However, this question will not be discussed in this paper, as this is not the focus of the paper.

Before discussing the main objectives of the paper, let us define what recursion is. Summarising definitions of recursion by several authors (Chomsky 1965; Hauser et al. 2002; Roeper 2007; Bisetto 2010; Corballis 2011; Ralli 2013), the author defines recursion as a phenomenon of cyclic fashion to create sentences, phrases or words, as complex or long as we like. Here, complex means embedding of phrases or words within phrases or words of the same kind. In principle, it is possible to construct limitless embedding structures in human language, at least, within the limitations of one's memory and processing capacity. At word level especially compounding is a word formation process where recursiveness is more widely attested.

Roeper (2007) gives examples at each level, phrase, clause and word levels. They are presented in the examples below (1)–(3).

(1) Phrase level
   a. Possessive: John's friend's car's motor
   b. Prepositional phrase: in the kitchen, in the cabinet, in the corner
   c. Conjunction: and I came and I saw and I conquered, John and Bill and Susan

(2) Clause level
   a. Infinitive: John wants to start to go to sing
   b. Finite: Mary thinks I think you think she did it

(3) Word level
   a. Prefixation: re-re-read, anti-anti-missile
   b. Adjective: big, black, strange bear
   c. Compound: student film group festival

(Roeper 2007: 108)

As (1)–(3) show, at phrase level, clause level and word level, it is possible to have embedding of phrases or words within phrases or words of the same kind. For example, (3c) shows that four nouns are merged together to create a word. In addition, unlike at phrase level, it is impossible to have another noun within this compound, or otherwise the resulting compound would be a completely different word. In other words, recursion of compounding obeys the Lexical Integrity Principle (Bresnan & Mchombo 1995). The Lexical Integrity Principle disallows insertion of an extra element inside. Thus, recursive compounds are more like words in this sense.

With the definition of recursion in mind, it is also important to define left-branching and right-branching recursive compounds. A left-branching recursive compound is one where a two-member compound is formed and another head is merged at the right-hand side, whereas right-branching is where a noun is added to modify a two-member compound. The resulting interpretation is different according to the branching direction (Roeper et al. 2002). A typical example of right-branching and left-branching recursive compound is as follows.

(4) student (*the) film committee [= 'student committee on the films']

(5) student film (*the) committee [= 'the committee on student films']

(Roeper et al. 2002: 2)

Roeper et al. argue that in both (4) and (5), overt functional heads, such as determiners, are disallowed. (4) is a right-branching recursive compound.[1] The two-member compound *film committee* is modified by *student*, whereas (5) is a left-branching recursive compound. The two-member compound *student film* is a modifier for the noun *committee.* The interpretation of the whole compound (4) and (5) is provided in the brackets.

According to Berg (2006: 197), compounding is a quite productive word-formation process in English, and its productivity extends not only to the semantic

[1] Bisetto (2010: 15) argues that recursion is said to be distinguished from iteration. What we call right-branching recursive compounds are not recursive compounds, because they are just repetition with the same head. However, in this paper, I argue that right-branching compounds are also recursive compounds, since at phrase level and clause level it is possible to have recursion of right-branching structure in many languages.

relationships holding between the constituents of compounds, but also to the number of compounds. The example (3c) given by Roeper (2007) is a representative example of this type of compounds. In this paper, this is assumed to be true, not only in English but also in Japanese. Some more examples of four-member compounds in English, some taken from Berg and others from my observations (7)–(9) in English and those (10)–(13) in Japanese.

(6) a. child language acquisition research
b. child language acquisition research group
c. child language acquisition research group member
(Berg, 2006: 198)

(7) timeshare property ownership[2]

(8) UN Security Council Resolutions

(9) muscle contraction mechanism calcium

(10) nihongo-kyouiku-nouryoku-kennteishiken
Japanese-teaching-competency-examination
'Japanese Language Teaching Competency Examination'

(11) omelet rice kaido project
omelet rice highway project
'Omelet-Rice Highway Project[3]'

(12) kakuheiki-haizetsu-kokusai-campaign
nuclear-abolishment-international-campaign
'Treaty on the Prohibition of Nuclear Weapons'

(13) Kikoh-hendoh-wakugumi-johyaku
Climate-change-framework-treaty
'United Nations Framework Convention on Climate Change'

The data above clearly shows that recursivity of compounds with more than four members is also observed in English and Japanese. However, this paper focuses on three-member compounds. Ambiguity of four-member compounds is in need of further research.

This paper is organised in the following way: the next section will explain a corpus study of the compounds with native speakers' judgements of each language. This section is followed by a discussion on some potential reasons for restriction of right-branching recursive compounds. In the conclusion, the findings of the studies will be summarised.

2. Corpus study of recursive compounds

As discussed in the previous section, right-branching recursive compounds are more restricted than left-branching recursive compounds in English and Japanese. In this

[2] This data is from *British National Corpus*.
[3] This is a name of a project conducted by a village, Hidaka, in Kochi Prefecture, Japan.

section, I will first present data in these languages from my observations. This is followed by the corpus study of recursive compounds with native speakers' judgements.

2.1 *Recursive compounds in English and Japanese*

Recursive compounds can be observed in languages across the world, and English and Japanese are not exceptions. They are usually constituted of three nouns and denote product names, names of organisations, descriptions of phenomena, restaurant menus. They are also productively produced by children. An alternative strategy for the speakers would be using the corresponding phrases. The following examples are taken from the literature and my observations in everyday life in English and Japanese.

(14) a. [[air traffic] control]
b. # [air [traffic control]]

(15) a. [[student film] group]
b. [student [film group]]

(16) a. [[peanut butter] sandwich][4]
b. # [peanut [butter sandwich]]

(17) a. [[toilet paper] tissue]
b. [toilet [paper tissue]][5]

(18) a. [[logpile] house]
b. *[log [pile house]]

(19) a. [[underground] house][6]
b. *[under [ground house]]

(20) a. [[potato chip] plant]
b. # [potato [chip plant]]

(21) a. [[student film] committee]
b. [student [film committee]]

(Roeper, et al. 2002: 2)

(22) a. [mail [delivery service]]
b. [[mail delivery] service]

(23) a. [rail [network]][7]
b. *[[rail net] work]

The examples (a) are left-branching recursive compounds, whereas (b) are the corresponding right-branching recursive compounds. The original examples found are left-branching recursive compounds in (14)–(21), and (22)–(23) are right-branching ones. The examples (14)–(20) are preferably interpreted as left-branching, not right-

[4] A menu in a café and Google UK.
[5] My observation – toilet.
[6] (19) and (20) are from the children's book *The Gruffalo* (Donaldson 1999)
[7] (22) and (23) are from Google UK.

branching, by native speakers. In addition, the lexicalised compounds such as *logpile* or *network* cannot be separated into two different words. These examples show that there are more left-branching recursive compounds than right-branching recursive compounds in English.

Let us observe some phonological characteristics of these recursive compounds. According to the Compound Stress Rule (Lieberman & Prince 1977), the branching-direction is responsible for stress assignment in NNN compounds. In left-branching recursive compounds, in (14a)–(21a) the left most constituent is assigned highest prominence whereas in (14b)–(21b), (22a) and (23b), the second constituent is the most prominent one. For example, (14a) the stress is placed on *air*, whereas in (22a), *delivery* has the most prominent stress.

In addition, the children's data from CHILDES are all left-branching recursive compounds, not right-branching recursive compounds. The following are some representative examples. From the interpretation, the brackets on each compound are added to show the branching direction.

(24) a. [[Christmas tree] cookie] = 'a cookie shaped of Christmas tree'
b. [[peanut butter] sandwich] = 'a sandwich with peanut butter'
c. [[baby doll] napkin] = 'a napkin for a baby doll'
d. [[nursery school] book] = 'a book at nursery schools'
(Roeper, et al. 2002: 3)

What about in Japanese? Left-branching recursive compounds are more productive than right-branching recursive compounds, like in English.

(25) a. [[doitsu bungaku] kyoukai]
Germany literature association
'association for German literature' (Nishiyama 2015: 79)
b. [doitsu [bungaku kyoukai]]
'literature association in Germany' (Nishiyama 2015: 79)

(26) a. [[nise danuki] shiru]
fake badger soup
'soup with fake badger (either chicken or fox)' (Tokizaki 2011: 5)
b. [nise [takuki ziru]]
'plastic model of badger soup (presented in restaurant for customers)' (Tokizaki 2011: 5)

(27) a. [[anpan man] myuuziamu]
Anpan man museum
'museum of a character Anpanman'[8]
b. *[anpan [man myuuziamu]]

(28) a. [[ushiro muki] kaidan]
back direct-FINITE stairs
'stairs for people going backward'[9]
b. # [ushiro [muki kaidan]]

[8] This museum is in Kochi, Japan.
[9] My son produced this, when he was 3 years old, walking backward on a staircase.

(29) a. [[kodomo hon] kurabu]
child book club
'club for children's book'
b. # 'book club for children' (cf. kodomo no hon kurabu, with GEN *no*)

(30) a. [bikini hibaku] sosho]
Bikini hibaku lawsuit
'lawsuit for survivor of hydrogen bomb test at Bikini atoll'[10]
b. #'hibakusha lawsuit for Bikini atoll'

The above examples of Japanese recursive compounds collected from observations and the literature on compounding also show that left-branching recursive compounds are more productive than right-branching recursive compounds, both in number and interpretation. Let us consider some characteristics of these recursive compounds. Note that the contrast between (25a) and (25b) is not only the interpretation, but also the accent. According to Nishiyama (2015: 79), (a) has word accent on the vowel of *kyo*, since the whole compound is phonologically one word. In contrast (b) has phrasal accents on both *do* and *kyo*. So, the right-branching phonologically behaves more like a phrase. Similarly, (26a) and (26b) have different accents.

Notice that in the example (26), when the interpretation is left-branching (26a), the initial sound of *tanuki* /t/ becomes voiced when combined with another word, whereas in (26b), the initial sound of *soup* /s/ becomes voiced. This is a typical example of RENDAKU, Japanese sequential voicing. This phenomenon appears in the initial sound of the second constituent of a compound, so sequential voicing is a sign of a compound word.[11]

In addition, it is possible to have the right-branching interpretation, represented in (29b), not only the left-branching, represented in (29a). However, it is better to have the genitive case marker between the first and second constituents. In summary, this sub-section has shown that right-branching are more restricted than left-branching recursive compounds from my own observations and the literature on compounding in English and Japanese. It is not impossible to have a right-branching recursive compound, but it is somehow more restricted, as show in this sub-section. In addition, some phonological characteristics of recursive compounds were observed.

2.2 *Corpus studies in English and Japanese*

2.2.1 *Previous studies*

Kösling & Plag (2009) conducted an empirical study of phonological characteristics to see if NNNs overall conform to the *Compound Stress Rule's* (Lieberman & Prince 1977) predictions or not (See §2.1 for more details). To test if this rule is right, an empirical study was conducted. When taking acoustic data from the Boston University Radio Speech acoustic data, out of 448 NNN compounds, 326 were classified as being left-branching, 122 as right-branching. Their aim of the study was not to see the difference of productivity of left-branching in contrast to right-branching recursive

[10] Politics/Economics, Kochi Newspaper, 21st July, 2018.

[11] Otsu (1980) argues that Rendaku applies only to elements that are on a right branch at the lowest level of compound structure (Right Branch Condition, Otsu 1980). For example, (26a), *[[nise tanuki] shiru]*, the word *tanuki* is at the lowest level of compound structure. Here, the initial sound /t/ becomes voiced according to the Right Branching Condition. In contrast, (26b), *[nise [tankuki shiru]] shiru* is at the lowest level of compound structure. So, the initial sound becomes voiced.

compounds. However, from their sampled data, it is possible to see left-branching recursive compounds are more frequent than right-branching.

Another study of triconstituent was conducted by Lauer (1995) and it was found that left-branching recursive compounds are twice as frequent right-branching recursive compounds in English. He also proposes that compounds with more than three nouns are syntactically ambiguous. From these studies, phonological characteristics and systematic computer analysis were conducted. However, the productivity of recursive compounds in English and Japanese has not been conducted. In addition, the interpretations are assumed to be correct and have not been empirically tested by native speakers of English. For these reasons, the present study needs to be conducted.

2.2.2 *The National Web Japanese Corpus and the Corpus of Contemporary American English*

The Japanese data used in this study is taken from the *National Web Japanese Corpus* (NWJC). The *National Institute for Japanese Language and Linguistics*, Japan (NINJAL) has compiled a ten-billion-word scale Japanese web corpus named *NINJAL Web Japanese Corpus* (hereafter NWJC) (Asahara et al. 2014).

The *National Web Japanese Corpus* was chosen for this study because of the following reasons. First, the corpus contains data from a number of resources, including the product names, names of organizations, descriptions of phenomena, restaurant menus, which we expected to contain a fair number of NNN compounds. Second, it is a representative corpus of Japanese. Another study of recursive compounds in Japanese was conducted by Koyama (2009). Using a Japanese domain corpus, they obtained hierarchical relations of recursive compounds. The present study explores interpretations and structures of recursive compounds in Japanese, more than specialised terms.

For the English data, the *Corpus of Contemporary American English* was used. 100 million words and which is freely available from Prof. Mark Davies's website, among other online services. The corpus was used for this study, because this is the largest corpus of American English and contains a number of samples, including news genres and many others (§2.1), which are expected to have recursive compounds. It was not possible to find tags for recursive compounds, so I asked the founder of the corpus, Prof. Mark Davies to get NNN strings and to rank them in order.

The 40 highest frequency words were the data for the experiment. The researcher conducted a control experiment for each language to see if the data collected were words that average educated speakers would be familiar with. For the first aim of this paper, the following experiment was conducted.

2.2.3 *Procedure*

For the experiment for English and Japanese recursive compounds, eight monolingual native speakers of English and Japanese, aged between 30 and 80, were asked to provide their intuitive interpretation for each target NNN string. For English, all the participants grew up in the UK and five of them had been living there all their lives and three had left the UK to live in another country and one in Japan for the last 5 years at the time of the experiment.[12] For Japanese, all grew up in Japan and three had left Japan

[12] The participants are British English native speakers, not American English. This is because the researcher has easier access to British native speakers. As discussed later, the place where the participants grew up or lived does not seem to affect their knowledge of the target NNN strings. However, for future research it would be interesting to see the difference in the interpretation between the regional differences of the participants.

to live in another country. The researcher conducted pre-study on the participants who had lived abroad, and it was clear that the place where they had lived did not affect their interpretation. However, it was made sure that they all had at least higher education and/either read newspapers regularly to have enough data for their interpretations. Before the experiment, the participants were asked to read the instructions provided to them on a sheet of paper. During the experiment they were given as long as they needed to write the interpretations after they had read the word. Depending on the subjects' thinking and writing speed, each session took between 15 and 25 minutes. They had time to ask questions to the researcher if necessary. In addition, they were asked to write down their immediate interpretation for each word when they first read the word. In total 320 interpretations for each language were collected (40 NNN strings x 8 native speakers).

After obtaining the data from the participants, the researcher observed their interpretations and paraphrasing carefully, and divided them into right-branching and left-branching interpretations. The intuition of the researcher was used for the categorisation. Right-branching is where a two-member compound is modified by a modifier, whereas left-branching is where a two-member compound modifies another noun. For example, for the NNN string of *vanilla ice cream*, the participants' interpretation was 'vanilla-flavoured ice cream', so it is a right-branching recursive compound. The compound *ice cream* is modified by *vanilla*.

In summary, left-branching recursive compounds and right-branching recursive compounds were observed. In addition, it was summarised that left-branching recursive compounds are more frequent than right-branching recursive compounds. Next, the details of the present study were discussed including the descriptions and reasons for using the two corpora. The next section will be the results and discussions of the study.

3. Results and Discussions

3.1 *Results of English native speakers' judgements*

In both English and Japanese, the forty NNN strings were tested for native speakers' judgements. Japanese left-branching recursive compounds are 9 times as productive (frequent) as right-branching in terms of the interpretation given by the participants. In English, left-branching recursive compounds are 3 times as productive as right-branching in terms of interpretation. Table 1 summarises this. 1 percent of the Japanese participants and 4 percent of the English ones said they did not know.

Table 1: Interpretations of the data (%)

	Japanese	**English**
Right-branching	11	20
Left-branching	88	76
I don't know	1	4

The first aim of the paper: the hypothesis that right-branching recursive compounds in Japanese and English are more restricted in frequency than left-branching is borne out. As discussed in §1, this is based on the assumption that Baayen & Lieber (1997) are right in that the higher the token frequency of word, the higher the productivity of the

word. The 40 highest frequency NNN strings from these corpora were interpreted as left-branching more frequently. To show this is true in Japanese, typical examples of Japanese with the native speakers' interpretations are shown below.

(31) [[omu raisu] bento]
omelette rice lunchbox
'lunch-box which only has omelette rice'

(32) [[zidoh hukushi] center]
children welfare center
'center for children's welfare'

(33) [[zinken hooritsu] sohdankai]
human-rights law advisory-meeting
'advisory meeting for human rights law'

(34) [[zidoo kurabu] setsumeikai]
child club orientation
'orientation for children's club'

(35) [[nihon bungaku] kyookai]
Japan literature association
'association for Japanese literature'

(36) [shinbun [yoron choosa]]
newspaper public-opinion research
'public opinion research organised by newspaper company'

All the eight participants interpreted the data in the examples as represented in the brackets. Two of the participants said they did not know two of the target strings. However, these data clearly show that the participants prefer interpretation of left-branching recursive compounds than that of right-branching recursive compounds. This is especially true when the NNN strings are ambiguous, like (34) and (35). For example, (34) can be interpreted as 'club orientation for children'. This string might be more easily interpreted as 'orientation for children's club', since the term *zido club* is used recently to describe an organisation for school children before/after school to be minded by city-employees. So, *zido club* is now a compound in native speakers' lexicon. Similarly, (35) can be interpreted 'literature association in Japan' as well as 'association for Japanese literature'. However, all the participants preferred the latter, the left-branching interpretation.

For English native speakers' judgements, typical examples are represented below.

(37) [health care] reform]]
'Health care reform can occur when government policies call for change in practices and ideas surrounding a nation's health system'

(38) [world health] organization]]
'an organization to help the health of the people of the world'

(39) [public opinion] poll]]
'A system of gathering information from the public that determines a national viewpoint on topics through local voting system'

(40) [reader service] card]]
'a notification or advertisement in a library'

(41) [[vanilla [ice cream]
'a dessert made from blending and freezing cream, milk, eggs, sugar, and vanilla'

The interpretation written underneath for each string is the same for all the participants, although some of the wordings are different. Two of the participants said they did not know and were not sure about the interpretations for two of the strings. Interestingly, in English, names of organisations are observed to have the preposition of in-between the constituents. This is more so than in Japanese with a genitive case marker. However, from the observations and the corpus, recursive compounds are not unproductive. In fact, it is more productive and observed, at least in the corpus used for this study.

It is important to note the following fact. In (41) it is easier to interpret *ice cream* as a compound and vanilla is its flavour than cream of vanilla-ice. There are some examples like this, where it is easier to have the interpretation that the participants wrote. However, the main point here is that the frequency of left-branching interpretation by the native speakers is higher than that of right-branching. So the researchers hypothesis for English is borne out.

3.2 *Potential reasons for restrictions*

In the previous sub-section, it was found that in both English and Japanese, there are more left-branching recursive compounds than right-branching recursive compounds with native speakers' judgements. Let us discuss some of factors that could be responsible for restricted recursion in right-branching recursive compounds in contrast to left-branching recursive compounds. The factors discussed by Pöll (2015) are represented in this paper.

First, he cites Haider (2001) who argues that the direction of headedness, i.e. left-headedness in Romance languages vs. right-headedness in Germanic languages is the reason behind productivity of recursive compounds.

(42) a. French: *[poisson [chat [bébé]]]
'lit. fish cat baby; baby cat fish' (Haider 2001: 1)
b. Italian: *[gatto [capo stazione]]
'lit. cat head station; stationmaster cat' (Pöll 2015: 9)

Haider (2001) and Pöll (2015) argue that in Romance languages, it is not possible to translate into English the counterpart of the data, fish cat baby, for example. In (42a) and (42b), the left-most constituent, *poisson* or *gatto*, is the head of the whole string. However, it is not possible to have these strings in these languages, and Haider argues this is due to the left-headedness of these languages. However, this reason, as also Pöll argues, is not valid when one looks at other languages with left-head compounds, like Vietnamese (Dang 2013: 21) or Malay (Post: personal communication). Pöll also argues that the above examples are in fact productive in the languages.

Another factor Pöll (2015) discusses is the existence of a linking element. This reason is argued by Mukai (2008). In Swedish, according to Josefsson (1997), a linking element needs to exist in left-branching recursive compounds, or otherwise the strings would be ungrammatical. Some typical examples are represented below.

(43) a. [barn [bogklubb]]
child book-club
'book club for children'
b. [[barnbok]-s-klubb]
children-book club
'club for children's book'

The example above shows the difference in the interpretation, and to have the left-branching interpretation, *club for children's book*, it is obligatory to have a linking element. With this in mind, Mukai (2008) argues that the existence of a linking element is the reason behind the productivity of left-branching recursive compounds in English, Japanese and Mainland Scandinavian. However, Pöll argues that this is not true, since there is no linking element in English or Japanese. In Norwegian or Danish, the element is not obligatory[13]. In addition, in Dutch or German where the linking element does exist in recursive compounds, it is not obligatory either. As a result, the existence of a linking element does not seem to be true for the reason, either.

In this paper, the fact that processing is the most possible reason for the restriction, at least in English and Japanese, is argued. According to Pöll, compounds of the type [[AB] C] are easier to parse, since a constituent can be formed earlier than in [A [BC]] structures. In left-branching compounds, the speaker forms a constituent out of adjacent roots earlier than in right-branching compounds, where a constituent cannot be formed until the last root is pronounced (Hawkins 1990; Sugioka 2008). For instance, upon hearing or speaking it is easier for the speaker/hearer to form the AB as a two-member compound (left-branching recursive compound).

In this section, from the study of the corpus with native speakers' judgements in Japanese and English, left-branching recursive compounds are more frequently interpreted than right-branching recursive compounds. Thus, the hypothesis is borne out for this empirical study. In addition, the author argued that Pöll's reason of processing is the most possible behind the restriction of right-branching recursive compounds in English and Japanese.

5. Conclusion

The aim of this paper has been to test the hypothesis that left-branching recursive compounds are more easily interpreted than right-branching recursive compounds in English and Japanese. To test the hypothesis, empirical studies with corpus and native speakers' judgements were conducted. From the studies, it was clear that the hypothesis is borne out, i.e. left-branching recursive compounds are more preferably interpreted than right-branching recursive compounds. This means that the two-member compound can be a modifier of a noun. It is possible to have another modifier for a two-member compound (right-branching), but this is less frequent. However, when presenting the data to native speakers for their judgements of interpretation, it is necessary to consider the influence of already-existing/lexicalized compounds (e.g. *vanilla ice cream*, where

[13] I would like to thank Professor Anders Holmberg for his comments on this subject.

ice cream is already a lexicalized compound in native speakers' minds).

In addition, some potential factors for the restrictions of right-branching recursive compounds have been discussed along with Poll's (2005) reasons, and the reason is that it is easier for the speaker/hearer to parse left-branching recursive compounds, just like Pöll argues.

However, it is necessary to consider the following facts to prove that the reason of parsing is right. First, more data needs to be collected with native speakers' judgements in English and Japanese. In addition, it is important to consider the difference of productivity of recursive compounds between different languages across the world.

Acknowledgements

I would like to thank Professor Mark Davies for obtaining the data from the corpus. I also would like to thank the eight native speakers of Japanese and English. All the errors are mine.

References

Asahara, Masahara & Maekawa, Kikuo & Imada, Mizuho & Kato, Sachi & Konishi, Hikari. 2014. Archiving and Analysing Techniques of the Ultra-large-scale Web-based Corpus Project of NINJAL, Japan. *Alexandria* 2(12). 129–148.

Baayen, Harald & Lieber, Rochelle. 1997. Word frequency distributions and Lexical Semantics. *Computers and Humanities* 30. 281–191.

Berg, Thomas. 2006. The internal structure of four-noun compounds in English and German. *Corpus Linguistics and Linguistic Theory* 2(2). 197–231.

Bisetto, Antonietta. 2010. Recursiveness and Italian compounds. *SKASE Journal of Theoretical Linguistics* 7(1). 7–35.

Bresnan, Joan & Mchomo, Sam. 1995. The Lexical Integrity Principle: Evidence from Bantu *Natural Language & Linguistic Theory* 13(2). 181–254

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

Corballis, Michael. 2011. *The Recursive Mind: The Origins of Human Language, Thought, and Civilization.* Princeton: Princeton University Press.

Dang, Zimmer. 2013. *Does prosodic word recursion cause phonetic initial strengthening?* California State University, Fresno. (MA dissertation.)

Donaldson, Julia. 1999. *The gruffalo*. Illustrated by Scheffler, Axel. Macmillan: Macmillan Children's books.

Everett, Daniel. 2005. Cultural constraints on Grammar and cognition in Pirahã. *Current Anthropology* 46(4). 621–646.

Haider, Hubert. 2001. Why are there no complex head-initial compounds? In Schauer-Wolles, Chris & Rennison, John & Neubarth, Friedrich (eds.), *Naturally!*, 165–174. Torino:

Rosenberg & Sellier.

Hauser, Marc D. & Chomsky, Noam & Fitch, W. Tecumseh. 2002. The Faculty of Language: What is it, Who has it, and How did it Evolve? *Science* 98.1569–1579.

Hawkins, John. A. 1990. Parsing theory of word order universals. *Linguistic Inquiry* 21. 223–261.

Josefsson, Gunlög. 1997. *On the Principles of Word Formation in Swedish*. Lund: Lund University Press.

Kosling, Kristina & Plag, Ingo. 2009. Does branching direction determine prominence assignment? An empirical investigation of triconstituent compounds in English. *Corpus Linguistics and Linguistic Theory* 5(2). 201–239.

Koyama, Teruo. 2009. Composite Term Extraction from Japanese Text. *Joho Chishiki Gakkaishi* 1(4). 306–315.

Lauer, Mark. 1995. Corpus statistics meet the noun compound: Some empirical results. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 47–54. Stroudsburg, PA: Association for Computational Linguistics.

Liberman, Mark & Prince, Alan. 1977. On stress and Linguistic Rhythm. *Linguistic Inquiry* 8(2). 249 –336.

Mukai, Makiko. 2008. Recursive compounds. *Word Structure* 1(2). 178–198.

Mukai, Makiko, 2017. Semantic characteristics of recursive compounds. In Sheehan, Michelle & Bailey, Laura R. (eds.), *Order and structure in syntax II: Subjecthood and argument structure*, 285–288. Berlin: Language Science Press.

Nishiyama, Kunio. 2015. Nihongo no kukozo (Structure of phrases in Japanese). In Nishihara, Tetsuo & Shinichi, Tanaka (eds.), *Gendai no keitairon to onseigaku・oninron no shiten to ronten* [Aspects and discussions in current morphology and phonetics-phonology], 78–95. Tokyo: Kaitakusha.

Otsu, Yukio. 1980. Some aspects of rendaku in Japanese and related problems. In Otsu, Yukio & Farmer, Anne (eds.), *Theoretical issues in Japanese linguistics* (MIT Working Papers in Linguistics 2), 202–278. Cambridge, MA: MIT Department of Linguistics and Philology.

Pöll, Bernhard. 2015. Restricted recursion in N-N compounding: some thoughts on possible reasons. *Linguistische Berichte* 242. 141–163.

Ralli, Angela. 2013. Compounds and its locus of realization: Evidence from Greek and Turkish. *Word Structure* 6(2). 181–200.

Roeper, Thomas. 2007. *The Prism of Grammar: How child language illuminates humanism.* Cambridge, MA: MIT Press.

Roeper, Thomas & Snyder, William & Hiramatsu, Kazuko. 2002. Learnability in a Minimalist Framework: Root Compounds, Merger, and the Syntax-Morphology Interface. In Lasser, Ingeborg (ed.), *The Process of Language Acquisition: Proceedings of the 1999 GALA Conference*, 25–35. Frankfurt: Peter Lang Verlag.

Sugioka, Yoko. 2008. Remarks on asymmetry and recursion in compound formation. In Sano, Tetsuya & Endo, Mika & Isobe, Miwa & Otaki, Koichi & Sugisaki, Koji & Suzuki, Takeru. (eds.), *An enterprise in the cognitive science of language: A festschrift for Yukio Otsu*, 65–78. Tokyo: Hituzi Shobo.

Tokizaki, Hisao. 2011. The Nature of Linear Information in the Morphosyntax-PF Interface. *English Linguistics* 28 (2). 227–257.

Makiko Mukai
Department of Cultural Studies
University of Kochi
2-22 Eikokuji
Kochi 780-8515
Japan
mukai@cc.u-kochi.ac.jp

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/03.pdf. ISSN 1336-782X.