

Ordering a Left-branching Language: Heaviness vs. Givenness

Hye-Won Choi*†

Ewha Womans University

Hye-Won Choi. 2009. Ordering a Left-branching Language: Heaviness vs. Givenness. *Language and Information* 13.1, 39–56. This paper investigates ordering alternation phenomena in Korean using the dative construction data from Sejong Corpus of Modern Korean (Kim, 2000). The paper first shows that syntactic weight and information structure are distinct and independent factors that influence word order in Korean. Moreover, it reveals that heaviness and givenness compete each other and exert diverging effects on word order, which contrasts the converging effects of these factors shown in word orders of right-branching languages like English. The typological variation of syntactic weight effect poses interesting theoretical and empirical questions, which are discussed in relation to processing efficiency in ordering. (Ewha Womans University)

Key words: Korean, word order, heaviness, syntactic weight, length, information structure, givenness, newness, corpus study, left-branching language

1. Introduction

Information structure and syntactic weight have been argued to be two major factors that influence constituent ordering in many languages (Collins, 1995; Givón, 1983; Givón, 1988; Gundel, 1988; Siewierska, 1993; Arnold et al., 2000; Dryer, 1980; Gibson, 1998; Gibson, 2000; Hawkins, 1994; Hawkins, 2004; Snyder, 2003; Wasow, 1997; Wasow, 2002). In a right-branching language like English, these two factors work toward the same direction of ordering: old or given information is placed before new information ('given before new' as in (1) below); a heavy or long constituent is placed after a light phrase ('short before long' as in (2)). As given information is usually shorter than new information (Givón, 1983), the two effects often converge

* A portion of this paper was presented at the 2009 Annual Conference of Linguistic Society of America (LSA) in San Francisco, January 2009. I thank the audience for their valuable questions and comments. I am also grateful to Joan Bresnan, Beth Levin, Ivan Sag, and Thomas Wasow for their insightful comments and discussions on various occasions. The final version of the paper benefited a lot from the comments by three anonymous reviewers. This research was supported by LG Yonam Foundation (Research Abroad Grant 2008).

† Department of English Language and Literature, Ewha Womans University, 11-1 Daehyun-dong Seodaemun-gu, Seoul 120-750, Korea. Email: hwchoi@ewha.ac.kr.

to push the short and given leftward and the long and new rightward. Because of this reason, some have argued that ordering phenomena can be explained by one of the two factors, not needing both, although which one of the two is responsible has been in controversy: some argued for syntactic weight (Hawkins, 1994), others for information structure (Siewierska, 1993; Snyder, 2003).

- (1) Given before new
 - a. Nixon's behavior gave Mailer an idea for a book.
 - b. *Nixon's behavior gave an idea for a book to Mailer.
- (2) Short before long
 - a. Nixon's behavior gave Mailer an idea for a book.
 - b. Nixon's behavior gave an idea for a book to every journalist living in New York city in the 1970s. (Snyder, 2003, 35)

Left-branching and head-final languages such as Korean and Japanese pose a challenge to explaining ordering phenomena because the factors mentioned above diverge in their directional effects. While given information is placed before new information just like in English and many other languages (Choi, 2008; Ferreira and Yoshita, 2003), a heavy constituent is placed before a light one in these languages (Choi, 2007; Gibson, 2000; Hawkins, 1994; Temperley, 2007; Yamashita and Chang, 2001), which is just the opposite to what is happening in right-branching languages like English. If we assume again that given information is usually light and new information is usually heavy, a heavy and new phrase, for instance, would get mixed signals: its heaviness would urge it to go to the left while its newness would push it rightward; similarly with a light and given phrase, to reverse directions. The two factors, if both of them indeed are in effect, would work not in a converging manner, but rather in a competing way, canceling out the effects of each other (see also Wasow (2002, 82)). This is probably why Hawkins (1994, 144) notes that "overall success rate . . . is lower" in left-branching languages with his syntactic-weight-based EIC theory.

By examining the ordering alternation of dative construction in Korean, this paper will investigate whether information structure and syntactic weight are indeed distinct and independent factors that influence word order in this left-branching language and if so, how they interact. Specifically, it will be examined whether the effects of information structure and syntactic weight diverge and compete each other. This paper is organized as follows. Section 2 displays basic facts about Korean dative construction based on the data collected from a million-word Sejong corpus. Section 3 analyzes the data according to the effects of information structure and syntactic weight. Section 4 discusses theoretical and empirical implications of the findings from section 3 and concludes the paper.

2. Data

2.1 Data collection: dative construction

For the current study of word order alternation, dative construction such as the one in (3) below is examined. Dative construction is defined to have two arguments besides the subject: one is a recipient argument marked by the dative marker *-ekey* or *-hanthey* ‘to’, and the other, a theme argument marked by the accusative marker *-(l)ul*. Analogous to the English counterparts (3a’) and (3b’), the order between the dative argument and the accusative argument can alternate, as shown in (3a) and (3b).

- (3) a. (Ku-ka) chinkwu-eykey chayk-ul cwu-ess-ta. [DA order]
 he-Nom friend-Dat book-Acc give-Pst-Dcl
- b. (Ku-ka) chayk-ul chinkwu-eykey cwu-ess-ta. [AD order]
 he-Nom book-Acc friend-Dat give-Pst-Dcl
- a’. He gave (his) friend a/the book.
- b’. He gave a/the book to (his) friend.

Dative construction is an ideal case for the study of ordering alternation because it allows us to examine ordering between two object-like arguments. In comparison, ordering between arguments and adjuncts has its limitations in that they inherently differ in their syntactic and pragmatic characteristics; for one thing, not all adjuncts can be associated with information structure. On the other hand, ordering between subject and object in transitive sentences can also be said to be already biased, given that subject has more prominent syntactic and discourse status: subject is usually the default topic and normally comes first in a sentence even in a free word order language like Korean. What makes it more difficult is that Korean is a *pro*-drop language where any argument—most frequently the subject—can be dropped as far as it is understandable and recoverable from the context. In fact, in the dataset collected for this study we find subject missing in more than half of the cases. And if subject is present, it is likely to be a non-omissible one probably due to its syntactic and/or pragmatic prominence. Therefore, one can assume that it would take stronger effects (may it be syntactic weight or information structure) to place an object before a non-omissible subject.¹

The current study is based on the dative construction data collected from a million-word, untagged “spoken” portion of Sejong Modern Korean Corpus (Kim, 2000), which consists of transcribed or pre-scripted texts of TV/radio drama shows, interviews, talk shows, and news shows as well as lectures and casual conversations. Clauses with both an overt dative argument (marked with *-ekey* or *-hanthey*) and an overt accusative argument (marked with *-(l)ul*) were collected, regardless of their matrix or subordinate status. Since Korean is a *pro*-drop language, there may

¹ Actually, the order between (overt) subject and object seems to be hardly alternating. In Hawkins’s (1994) data of Korean, the ratio of SOV to OSV is 176 to 6 (of the total 182 sentences).

be clauses where either dative or accusative argument (or even both) is missing, but those were disregarded (i.e. not collected) in this study because what we are interested in is the relative order between the two arguments.

The total of 714 tokens of dative construction were collected. While all six orders, shown in (4), are theoretically possible, none in the dataset are in (4e) or (4f) order, where both dative and accusative arguments are placed before subject. There are very few tokens in (4c) or (4d) order too, 7 and 4 respectively (1.5%), where one argument precedes the subject. Almost all clauses (98.5%) are in (4a) or (4b) order, where subject may or may not be present. As non-initial subjects are very rare and many subjects are missing, we will be only concerned with the relative position between the dative and the accusative arguments.

- (4) a. (Sbj) Dat Acc V
 b. (Sbj) Acc Dat V
 c. Dat Sbj Acc V
 d. Acc Sbj Dat V
 e. Dat Acc Sbj V
 f. Acc Dat Sbj V

We coded a clause with the dative argument preceding the accusative argument (regardless of the presence or position of the subject) as 'DA (dative before accusative)' order and one with the accusative argument preceding the dative argument as 'AD (accusative before dative)' order in this study. We have 599 tokens in the DA order and 115 in the AD order.²

(5)

Total	DA order	AD order
714	599	115
(100%)	(83.9%)	(16.1%)

2.2 Data coding

2.2.1 Marking information status. Information status of an argument, i.e., givenness or newness, is known to be a main factor that influences word order from early on (Givón, 1983; Givón, 1988; Gundel, 1988; Siewierska, 1993; Snyder, 2003, citing only a few).³ As first proposed by Givón (1983; 1988), the referential distance

² The ratio of DA-order tokens to AD ones in our data is 5.2 to 1, which is more skewed than what Hawkins (1994) reports in his study, which is 3.8 to 1, although the total number of dative tokens is only 24 in his dataset.

³ There is another information structural factor that can potentially influence word order: 'urgency' or 'importance' (Givón, 1983; Givón, 1988; Gundel, 1988). This factor is supposed to pressure a phrase leftward as well. As 'urgent' or 'important' information may well be 'new', this factor may be in conflict with the givenness factor. This conflicting nature of two informational factors is actually a main reason that Hawkins (1994) rejects pragmatic explanation of word order. However, conflicting factors are not really a problem and can be easily taken care of with recent developments of linguistic theories such as Optimality Theory or probabilistic or statistical modeling theories. A real problem lies in how to measure 'urgency' or 'importance' reliably. Due to this technical difficulty, this factor was not taken into consideration in this study but it still is something worth studying in future research.

(RD) of the argument was used as the standard to determine its givenness. If the current argument has an identical lexical item or coreferential expression (which may not be identical in form) within the 20 previous sentences, i.e., if the referential distance of the current argument to its coreferential expression is shorter than 20, it was marked ‘given’; if the current argument was not mentioned at all or its referential distance to the coreferential expression is longer than 20, it was marked ‘new’ (see Choi (2008) for details).

Interestingly, dative and accusative arguments in our dataset show different distribution in terms of givenness. While slightly more dative arguments are given (53.6%) than new (46.4%), a lot more accusative arguments are new (77.2%) than given (22.8%), as summarized in (6).

(6)		Dat		Acc		Total
	Given	383	(53.6%)	163	(22.8%)	546
	New	331	(46.4%)	551	(77.2%)	882
	Total	714	(100.0%)	714	(100.0%)	1428

This rather biased distribution in givenness may be one of the reasons why there are a lot more DA-order tokens than AD-order ones, if givenness indeed plays a role such that given information precedes new information.

2.2.2 Measuring syntactic weight. There may be several ways to measure syntactic weight. One of the main concerns is whether word order is more sensitive to syntactic complexity of a phrase or to the simple length of it, while complexity and length have proven to be correlated (Wasow, 2002). The number of words (Bresnan et al., 2007; Bresnan and Hay, 2008), the number of words that introduce new discourse referents (Gibson, 1998; Gibson, 2000), the number of nodes (Hawkins, 1994), that of phrasal nodes (Rickford et al., 1995), and finally the number of phonological phrases (Zec and Inkelas, 1990) have been used to measure weight. Although many corpus studies adopt the number of words within a phrase as the measure of its weight (length or complexity), this study takes the number of syllables as the measure (see also Choi (2007)). The main reason is, besides the fact that word length can vary a great deal from one syllable to a great many syllables, that it is hard to decide what counts as a word in Korean due to complex morphology whose boundaries do not necessarily coincide with semantic or syntactic boundaries: for instance, Hawkins (1994) counts case markers and complementizer-like verbal endings as separate words, which have traditionally been regarded as bound-morpheme suffixes. The number of syllables, on the contrary, is much easier to count thanks to the syllable-based orthography of Korean: conveniently, each orthographic unit equals to a syllable.

Measured in terms of the number of syllables, the length of dative arguments in the corpus dataset ranges from 1 to 48 syllables, and that of accusative arguments ranges from 1 to 59 syllables. Yet the mean length of dative argument is not significantly different from that of the accusative one (5.06 vs. 5.15). Also, the clauses where the dative argument is longer than the accusative argument comprise 24.1% of all the tokens, while those with longer accusative arguments are 24.8%.

So we can safely say that the data are not pre-biased in terms of argument length as to lean toward one order or the other.

3. Analysis

The current corpus study shows that Korean displays the ‘given before new’ ordering tendency like English, but ‘long before short’ preference unlike English.

3.1 Given before new

The clauses where given information precedes new information can be easily found in the corpus data. Two examples are illustrated below. (The context preceding the current sentence was presented in italics, translated in English.)

- (7) *Hwangja confessed to me yesterday. That she has a man.*

So . . .

ku namca-eykey kkohtapal-ul senmwulhalako yaykihaysseyo.

the man-Dat flowers-Acc present told

‘(I) told (her) to give the man flowers’

- (8) A: *When you focus on practicing medicine, you could get a little lax on your research. Or you could get a little slack in your practice and get complaints from patients.*

B: *Swamped by crowding patients, I often end up providing idea and money and . . .*

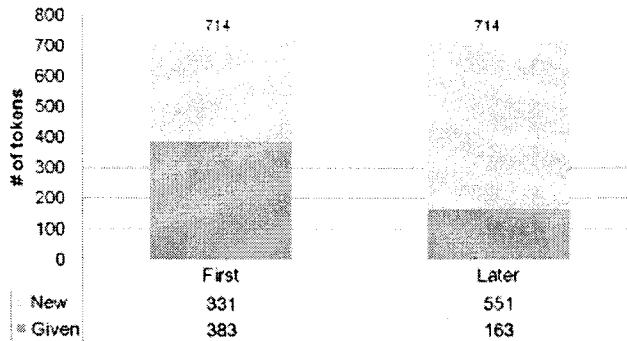
yenkwu-lul hwupaytul-eykey mathkinun swuto issko

research-Acc juniors-Dat leave

‘(I end up) leaving my research to junior doctors’

In (7), the dative argument *ku namca* ‘the man’ is given information (having its coreferential expression ‘a man’ (underlined) in the immediately preceding sentence), and this dative argument precedes the new-information accusative argument *kkohtapal* ‘a bunch of flowers’. In (8), on the other hand, the accusative argument *yenkwu* ‘research’ is given information (having its coreferential expression two sentences before) and precedes the new-information dative argument *hwupaytul* ‘junior (doctors)’.

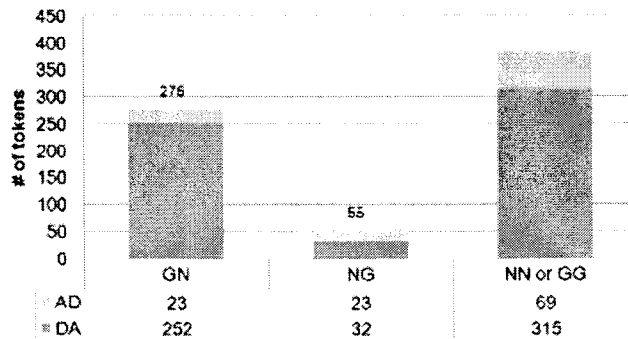
(9) Given vs. New



Over all, the corpus data prove that given information (whether it is dative or accusative) is more likely to come first than later between the two (Given first 70.1%; Given later 29.9%), while new information is more likely to come later than first (New later 62.5%; New first 37.5%). See the chart in (9) (given information is presented in dark grey and new information in light grey in the chart).

The chart in (9) shows the general trend that given information tends to come earlier in the sentence, whereas new information tends to be placed later. We can go further and see if given information is in fact followed by new information or vice versa. The following chart in (10) displays this distribution. GN signifies ‘given before new’ order and NG ‘new before given’. For instance, examples (7) and (8) are both of GN pattern as a ‘given’ argument is followed by a ‘new’ argument. NN or GG are those where arguments do not differ in their information status: ‘new before new’ or ‘given before given’ (see Choi (2008) for details).

(10) GN vs. NG



Putting aside those tokens where dative and accusative arguments do not differ in their information status (NN or GG), we see a great deal more GN (‘given before new’) tokens than NG (‘new before given’) tokens, the ratio being 5 to 1, as illustrated in (10). In other words, Korean shows strong ‘given before new’ preference in word order.

Now, let us look at the distribution more closely by breaking it down to DA order vs. AD order. The ratio of GN to NG in DA order is almost 8:1, but the

ratio of GN to NG in AD order is 1:1. In other words, a given dative argument is very likely to precede a new accusative, but a given accusative does not necessarily precede a new dative. We see a strong givenness effect in DA order but not in AD order. This may be related to the fact that there are a lot more new accusative arguments than given accusatives in the data. It may also have to do with the syntactic weight (length) effect, which will be discussed later.

3.2 Long before short

Unlike English, which shows the end weight effect as in heavy NP shift, a longer argument indeed tends to come before a shorter one in Korean. See examples below.

- (11) *llcali-lul ilhko noswukhanun silcikatul-eykey*
 [jobs-Acc lose and sleep on the street unemployed]-Dat
camcali-lul malyenhaycwuko
 shelter-Acc provide
 '(that) provides the unemployed who have lost their jobs and live on the street a shelter . . .'
- (12) *Cengpwu-nun*
 government-Top
cheycey nay-ey pwuphay-ka issesenun an toyntanun cem-ul
 [regime inside-in corruption-Nom exist must not point]-Acc
kwukmin cenchay-eykey kyoyuksikhiko issta.
 people all-Dat educate is
 'The government is educating the entire nation (about) the point that there must not be corruption within the regime.'

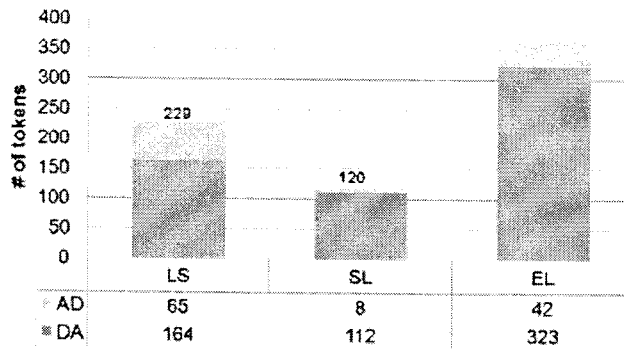
In (11), the long dative argument (14 syllables long except the 2-syllable-long dative marker, 'the unemployed people who have lost their jobs and live on the street'), precedes the much shorter accusative argument 'a shelter' (3 syllables long except the 1-syllable-long accusative marker). In (12), on the other hand, the long accusative argument 'the point that there must not be corruption within the regime' (16 syllables long) precedes the shorter dative argument 'the entire nation' (4 syllables long).⁴

Our corpus data indicate that LS ('long before short') tokens are almost twice as many as SL ('short before long') ones, while half the data are in EL ('equal length', with the difference being 0 to 2 syllables), as illustrated below.⁵ Korean does seem to have the 'long before short' ordering preference.

⁴ Not all preceding long arguments are syntactically complex in the dataset although the ones shown in (11) and (12) are.

⁵ Overall, less than a third (only 32%, 229 LS tokens) of the data are explained by the length effect. While length is indeed a strong factor, it seems quite stretchy to claim that syntactic weight alone can account for word order alternation (cf. Hawkins (1994)).

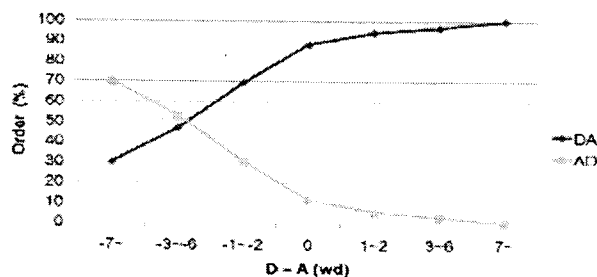
(13) LS vs. SL



Now, if we break it down to DA order vs. AD order, the ratio of LS to SL in DA order is 1.5 to 1 but the ratio of LS to SL in AD order is about 8 to 1. In other words, a longer accusative argument is much more likely to precede a shorter dative than a longer dative is likely to precede a shorter accusative. Length effect seems to be stronger with AD order. This is opposite to what we have seen in the information structure effect: the information structure effect is stronger with DA order.

To see how exactly length affects ordering, the proportions of DA and AD orders in percentage are plotted over the length difference between the dative and accusative arguments (length of dative minus length of accusative) in (14). As mentioned above, length was measured in syllables, but here it was transformed to the approximate number of “words” at the ratio of 3 syllables to 1 word (note that a “word” here is only a proxy, not a real one). Difference up to 2 syllables is regarded as no difference, so 0 word difference actually means that the difference is -2 to 2 syllables, and 1 word difference equals to the difference of 3 to 5 syllables, etc. When the difference is positive, it means that the dative argument is longer than the accusative argument, and when the difference is negative, the accusative argument is longer.

(14) DA vs. AD over length difference



The graph shows the general trend that as the accusative argument becomes longer (D-A being negative), the AD order (in grey) increases (and the DA order (in

black) decreases), and as the dative gets longer (D-A being positive), the AD order decreases and the DA order increases. As we can see on the left side of the graph, when the accusative argument is longer, especially when the difference is 3 words or bigger, the rate of AD order is in fact greater than the DA order, despite the highly skewed overall proportion: recall that the DA order is 83.9% of all the tokens and the AD is 16.1%. This is roughly the proportion when the difference is around 0.

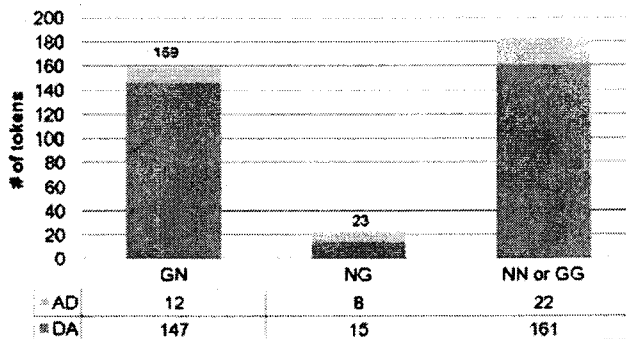
In other words, the fact that we see more AD-order tokens when the accusative is longer and more DA-order tokens when the dative is longer proves that longer argument is placed earlier in Korean and thus testifies the 'long before short' ordering preference.

3.3 When one factor is controlled

The fact that the syntactic weight effect is stronger in AD order but weaker in DA order and that the information structure effect is stronger in DA order but weaker in AD order seems to suggest that these two factors interact and one factor might actually weaken the effect of the other factor. One way to test this is to see what happens when the other factor is controlled.

First, we can see how information structure effect plays out when the length factor is controlled. Recall that half the tokens (365 tokens) show no difference in argument length. These are the ones where we can see the information structure effect without interference from the length effect. Compare the graph in (15) with the one in (10).

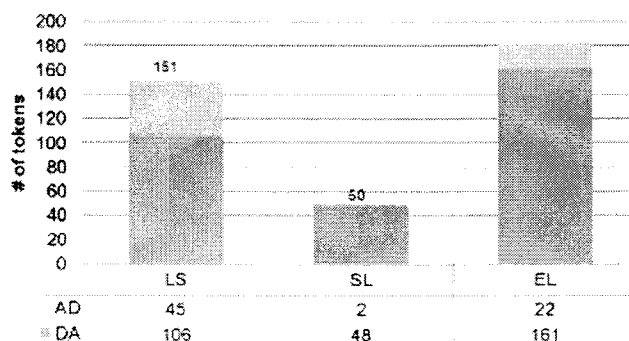
(15) GN vs. NG in EL



In these equal-length tokens, the information structure effect indeed shows more evidently. The ratio of GN (given before new) to NG (new before given) is almost 7:1, as opposed to overall 5:1, which we saw in (10). The information structure effect is stronger in each order too. In both DA and AD orders, the ratio of GN to NG is greater: in DA, 9.8 to 1 (as opposed to overall 8:1) and AD, 1.5 to 1 (as opposed to overall 1:1). That is, when the length effect is controlled, the information structure effect emerges in the AD order too. In other words, it appears that the strong length effect in AD order shadows the information structure effect in this order.

Now, by contrast, let's see what happens to the length effect when the information structure factor is controlled. We will see how the length effect plays out when there is no informational difference between the dative argument and the accusative argument, namely, when both are new (NN) or both are given (GG) (total of 384 tokens). As expected, the length effect is stronger, namely, the ratio of LS (long before short) to SL (short before long) is greater-3:1, as illustrated in (16), compared to 2:1 overall as shown in (13).

(16) LS vs. SL in NN or GG



The length effect is stronger in each order too. In both DA-order and AD-order tokens, the ratio of LS to SL is greater: in DA, 2.2 to 1 (compared to 1.5 to 1 in (13)) and AD, 23 to 1 (compared to 8 to 1). Again, we can see that when the information structure factor is silenced, the length effect emerges even stronger.

3.4 Information structure and syntactic weight together

We have seen above that the information structure effect and the syntactic weight effect in Korean do seem to interact and weaken the effect of each other in that one effect emerges stronger when the other effect is controlled. Now, let us investigate their interaction further by examining how the information structure effect is influenced depending on whether the syntactic weight plays a positive role or a negative one, namely, whether long comes before short or short comes before long. See the table in (17) below.

Let us look at the DA-order tokens first. Recall that the ratio of GN to NG in the DA tokens was 8 to 1, as shown in (10), and that the ratio grew bigger to 9.8 to 1 when the length effect was controlled, as illustrated in (15) and also repeated as the case of 'D-A=0' in (17). Now consider what happens if there is a positive length effect ('long before short'), that is, if dative is longer than accusative (D-A>0). The ratio of GN to NG in this case is about 3 to 1, which is lower than the average ratio, 8 to 1. By contrast, if there is a negative length effect ('short before long'), namely, if accusative is longer than dative (D-A<0), the ratio drastically increases to 20 to 1. In short, the syntactic weight effect works against the information structure effect. When the length factor is in effect (long before short), the information structure effect decreases; when the length factor is negatively affecting (short before long), the information structure effect increases.

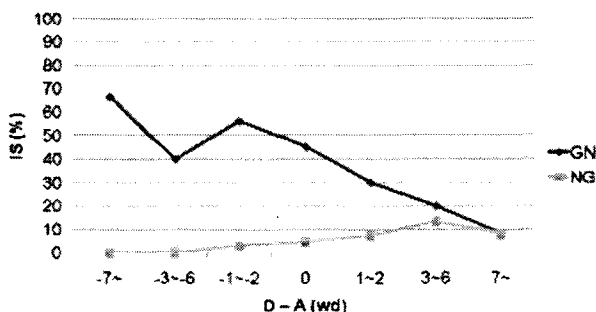
(17)

	DA order					AD order					Tot
	Sub Tot	GN	GG	NN	NG	Sub Tot	GN	GG	NN	NG	
D-A > 0	164 (100)	44 (26.9)	13 (7.9)	93 (56.7)	14 (8.5)	8 (100)	6 (75.0)	2 (25.0)	0 (0.0)	0 (0.0)	172
D-A = 0	323 (100)	147 (45.5)	49 (15.1)	112 (34.7)	15 (4.7)	42 (100)	12 (28.6)	14 (33.4)	8 (19.0)	8 (19.0)	365
D-A < 0	112 (100)	61 (54.5)	14 (12.5)	34 (30.4)	3 (2.6)	65 (100)	5 (7.7)	16 (24.6)	29 (44.6)	15 (23.1)	177
Tot (%)	599 (100)	252 (42.1)	76 (12.7)	239 (39.9)	32 (5.3)	115 (100)	23 (20.0)	32 (27.8)	37 (32.2)	23 (20.0)	714

We have seen above that the information structure effect was not very strong among the AD-order tokens. Yet, we see the same kind of interaction between the information structure effect and the syntactic weight effect in this order too. Remember that the average ratio of GN to NG was 1 to 1 but it was raised to 1.5 to 1 when there was no length difference ('D-A=0' in (17)). This ratio becomes even greater if the syntactic weight effect is negative ('short before long'), that is, if the accusative is shorter than the dative (D-A>0): 75% to 0%. This is exactly where we see the biggest information structure effect. By contrast, when the syntactic weight effect is positive (long before short), i.e., when the accusative is longer than dative (D-A<0), the ratio of GN to NG is actually reversed to 1 to 3: the information structure effect not only is weakened but actually disappears. Here again, we witness the contrary interaction between the information structure effect and the syntactic weight effect.

If we plot the proportions of GN and NG tokens over the length difference between the dative and accusative arguments, we can see the negative correlation between the two effects more clearly. See the graphs in (18) and (19).

(18) GN vs. NG over length difference in DA order

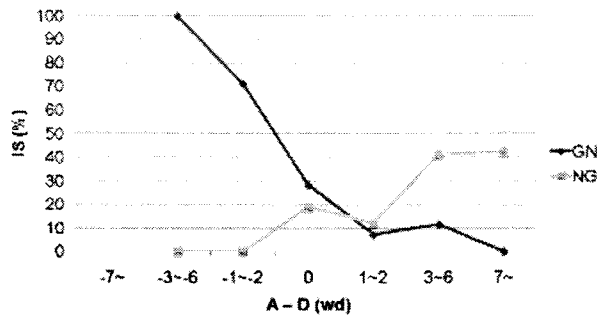


The left side of each graph is where the first argument (dative in (18) and accusative in (19)) is shorter than the second argument (accusative in (18) and dative in (19)), the mid point is where the first argument is of the same length of

the second argument, and the right side is where the first argument is longer than the second argument. In both the DA order in (18) and the AD order in (19), the information structure effect (represented by the GN line) decreases as the syntactic weight effect increases (i.e. as the first argument becomes longer). In other words, the contribution of information structure by GN to order decreases as the contribution of length increases. On the other hand, the NG line increases along with the increase of length difference, which means that the anti-information-structure effect is correlated with the length effect. This also shows that information structure and syntactic weight are in competition.

What differs between the DA order and the AD order is that the magnitude of the change in the information structure effect is a lot greater in the AD order: the GN line drops dramatically and the NG line soars as the length effect increases, as shown in (19), as opposed to the DA cases where the changes are milder, as shown in (18). Particularly, the negative contribution of information structure by NG becomes drastically greater as the length effect becomes evident in AD order. This is probably why the information structure effect is not so strong in AD order, namely the combination of strong length effect and negative information structure effect.

(19) GN vs. NG over length difference in AD order



To summarize, heaviness and givenness seem to compete in their influence on word order in Korean, so one effect becomes stronger when the other effect is controlled. In both DA and AD orders, the positive information structure effect represented by GN decreases and the negative information structure effect represented by NG increases as the first argument becomes longer, namely as the length effect becomes evident.

4. Discussion and Conclusion

As seen above, heaviness and givenness do seem to compete in their influence on word order in Korean and that one effect becomes stronger when the other effect is weakened and vice versa. After all, only 38.5% of our corpus data (275 tokens) are ordered given before new (GN) and 32.1% (229 tokens) are long before short (LS): neither factor alone can account for even half the tokens. ‘Given before new’ and ‘long before short’ are both valid ordering principles in Korean: each factor

explains a significant portion of the data especially when the other factor does not exert its power, while their effects get diminished when they clash.

This result is a strong argument against the claim that syntactic weight and information structure are not distinct factors and thus can be reduced to a single factor. As demonstrated in this paper, syntactic weight effect shows typological variation in its direction (Hawkins, 1994; Gibson, 2000; Yamashita and Chang, 2001; Temperley, 2007; Choi, 2007) and thus suggests its independence from information structure effect, which does not seem to show such variation. As a matter of fact, with recent developments in statistical methodology, quite a few researches have revealed that syntactic weight and information structure are clearly distinct factors that influence ordering alternation independently even in right-branching languages where the two effects converge (Arnold et al., 2000; Bresnan et al., 2007; Bresnan and Hay, 2008; Wasow, 2002).

Now, the question *why* naturally arises. Why do such factors as givenness and heaviness influence word order? One explanation is 'harmonic alignment', which refers to the phenomenon that more 'prominent' information is aligned with a more 'prominent' syntactic position (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2009). Citing a variety of psycholinguistic researches, Bresnan and Ford (2009, 10) argue that the prominence of information comes from 'activation,' which is "increased by lexical frequency, discourse accessibility, animacy, and effects of prior processing." Harmonic alignment is achieved such that "[t]he more activated units in the abstract cognitive representation of the message being formulated are expressed earlier in the incremental process of linearizing the sentence structure." This explains that given information, being more discourse accessible, is more activated than new information and thus comes earlier in the linear order of a sentence. Ferreira and Yoshita (2003) also argue that given information is more 'accessible' in processing and thus gets used earlier in forming a sentence than new information.

This processing-based explanation using the notion of activation or accessibility can also be applied to the syntactic weight effect of 'short before long' tendency in languages like English: namely, a short phrase—being more likely to be given and easier to process—is more accessible than a long phrase and thus comes earlier in the sentence (Wasow 2007). However, this explanation immediately faces a problem in accounting for the 'long before short' effect in Korean. If a short phrase is more accessible, then why would Korean put a long, less accessible phrase first, making processing (or production) more difficult? Alternatively, to maintain the tie between accessibility and processing, one would have to say that a long phrase is somehow more accessible in Korean. In fact, this is exactly what Yamashita and Chang (2001) argue about Japanese, which also exhibits the 'long before short' ordering preference. They argue that a long phrase is semantically richer and hence more accessible at the conceptual level (message-forming level), whereas a short phrase is more accessible at the form level (surface linearization level). They further argue that in Japanese, its syntactic freedom (i.e., free word order) does not hinder the conceptual preference of putting semantically richer, long phrases before short ones; however, in English, more rigid syntactic restrictions reduce the influence of conceptual factors, allowing form-related factors to play a larger role so that

formally more accessible, short phrases are placed before long ones.

Another explanation proposed for the ‘long before short’ tendency is the idea that languages tend to avoid a long constituent in the middle of a sentence, which makes processing difficult, and thus to rather have it either at the end (which is most preferred) or the beginning of the sentence (Dryer, 1980; Wasow, 2007). That is, the long and complex is preferred to go to the periphery of a sentence. A similar idea has been proposed to account for the cross-linguistically attested tendency to avoid center embedding. Specifically, Wasow (2007) argues that generally, a long constituent, which is more difficult for processing, tends to be postponed till the end of the sentence. However, in Korean or Japanese, where the verb is always in the final position, a long constituent cannot go to the sentence-final position; therefore, it goes to the beginning of the sentence. Unlike Yamashita and Chang’s, this approach maintains that long and complex phrases are consistently harder to process and thus to be postponed but that the language-specific syntactic property of Korean or Japanese leave only the second best position, i.e. sentence-initial position, available.

Finally, others have attempted to explain the ‘long before short’ effect in connection with the syntactic property of head directionality. The idea is that the syntactic head and its dependents must be close together; if not, it will incur bigger processing cost. This more or less same idea is shared by Hawkins’s (1994) EIC (Early Immediate Constituent) theory, Gibson’s (1998, 2000) DLT (Dependency Locality Theory) and Temperley’s (2007) DLMR (Dependency Locality Minimization Rule) although details may vary. In head-initial languages like English, the distance between the verb and its dependents (i.e., arguments) is shortest if the longer argument goes to the final position (assuming that the argument is recognized by its head (determiner), which is also phrase-initial); on the other hand, in head-final languages like Korean, the dependency distance (between the dependent arguments and the sentence-final verb) is shortest if the longer argument goes to the beginning of the sentence (assuming the argument is recognized by its head (case-marker), which is phrase-final). This is illustrated in (20) below.

- (20) a. John-i Mary-eykey [Suni-ka chayk-ul sasstanun kes-ul] malhayssta.
 John-Nom Mary-Dat [that Suni bought a book]-Acc told
 Hawkins’ CRD ratio: 39% (S 2/7, VP 3/6)
 Gibson’s Integration cost: 7 (Sbj 4; Dat 3; Acc 0)
- b. John-i [Suni-ka chayk-ul sasstanun kes-ul] Mary-eykey malhayssta.
 Hawkins’ CRD ratio: 64% (S 2/7, VP 3/3)
 Gibson’s Integration cost: 5 (Sbj 4; Dat 0; Acc 1)
- c. [Suni-ka chayk-ul sasstanun kes-ul] John-i Mary-eykey malhayssta.
 Hawkins’ CRD ratio: 87.5% (S 3/4, VP 2/2)
 Gibson’s Integration cost: 3 (Sbj 1; Dat 0; Acc 2)

In both Hawkins’s and Gibson’s theories (also Temperley’s, which is equivalent to Gibson’s in essence), the order in (20c), where the long accusative argument is sentence-initial, is the best order. Here the total distance between the verb and all

its arguments (underlined above) is the shortest: the aggregate node-to-word ratio is the biggest (Hawkins) and the integration cost is the smallest (Gibson).

While all three types of explanations briefly described above capture the 'long before short' preference in Korean, each differs in its empirical and theoretical implications. First of all, the accessibility-based theory by Yamashita and Chang does not assume any necessary connection between the 'long before short' tendency and the structural property of head-finality; rather the 'long before short' preference is related with word order freedom. A language with word order freedom would be freer to utilize the semantic/conceptual accessibility for word order and thus put a semantically richer, long constituent earlier without being restrained by the syntactic/form accessibility. According to their theory, therefore, we should be able to find a language that 1) has word order freedom, 2) is head-initial, 3) and shows the 'long before short ordering' tendency. Interestingly, this is exactly what the syntactic-dependency-based explanation (by Hawkins, Gibson, Temperley) predicts not to happen. Because head finality is inevitably related to the 'long before short' tendency, their accounts would not allow a head-initial language to exhibit the 'long before short' preference, regardless of word order freedom. On the other hand, the long-to-periphery account by Wasow generally prefers the 'short before long' ordering regardless of head directionality: theoretically, it would allow the 'short before long' ordering preference for a head-final language too if the verb is not rigidly sentence-final. As such, the three explanations make different predictions about the relationship between head-finality and syntactic weight effect. Therefore, the choice among these three approaches is certainly an empirical call and the Korean case alone cannot answer which is the best explanation. I will leave the answer open for future studies.

Another point of interest has to do with the position of the subject. As seen in (20) above, the syntactic-dependency-based theories by Hawkins, Gibson, or Temperley would all prefer the order in (20c), where the long accusative argument is at the beginning of the sentence, i.e., before the subject. The long-to-periphery account by Dryer or Wasow would also predict the long accusative argument to go to the beginning of the sentence since the final position is not available. However, our corpus data show that a long accusative argument does not always go all the way to the beginning of the sentence, but more often stays behind the subject, just like in (20b). The long accusative argument still precedes the shorter dative argument, but is very reluctant to precede the subject. We do have cases where the long accusative argument is at the beginning of the sentence but most of these are the ones where the subject is missing.⁶ As a matter of fact, the experiment data of Japanese presented by Yamashita and Chang (2001) also show that the long accusative argument tend to end up between the sentence-initial subject and the shorter dative argument rather than preceding both the subject and the dative

⁶ Recall that the corpus data used for this research was not coded for subject position but that only the relative position between dative and accusative arguments (DA or AD) were considered. Yet, among the AD tokens where the accusative is substantially longer (i.e., more than 9 syllables longer) than the dative, only one third of them have the accusative argument at the beginning of the clause; two thirds are the ones where either the subject or some other element precedes the accusative argument.

argument.

This suggests, first of all, that syntactic weight alone cannot provide full explanation of ordering phenomena. Subject comes first in many cases even if it is not the longest and even the longest object does not necessarily precede the subject. Whatever makes the subject special is certainly not its length. The information structural principle of ‘given before new’ obviously is a potential candidate for explanation, as subject is usually ‘given’ information and ‘givenness’ explains a significant portion of ordering between dative and accusative arguments. Yet, simple discourse-givenness, i.e., whether or not being mentioned previously, may not be enough; most of all, a subject that is mentioned before tends to be dropped in Korean. Therefore, a subject that is not dropped must have some other property than ‘givenness’ to survive, probably some sort of “prominence”, may it be pragmatic or syntactic. Although the ordering in relation to the subject is not the main topic of in this research, it certainly needs to be studied in depth for a full account of word order.

Setting the unresolved issues aside for future research, we conclude that no single factor can explain the ordering phenomena in Korean but rather multiple factors such as syntactic weight and information structure of givenness and probably more are involved in word ordering. Moreover, whatever is the underlying driving force of the ordering factors of syntactic weight and information structure may not come from a single source either. Whether it is cognitive activation/accessibility or processing efficiency, what explains one seems to counter the other. ‘Given before new’ and ‘long before short’ really look ununitable.

<References>

- Arnold, Jennifer, Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. Heaviness vs. Newness: The Effects of Complexity and Information Structure on Constituent Ordering. *Language* 76, 28–55.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Herald Baayen. 2007. Predicting the Dative Alternation. In G. Boume, I. Kraemer, and J. Zwarts (eds.), *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science, Amsterdam, pp. 69–94.
- Bresnan, Joan and Marilyn Ford. 2009. *Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English*. Stanford University and Griffith University. ms.
- Bresnan, Joan and Jennifer Hay. 2008. Gradient Grammar: An Effect of Animacy on the Syntax of *give* in New Zealand and American English. *Lingua* 118, 245–259.
- Choi, Hye-Won. 2007. Length and Order: A Corpus Study of Korean Dative-Accusative Construction. *Discourse and Cognition* 14, 207–227.
- Choi, Hye-Won. 2008. Beyond Grammatical Weight: A Corpus Study of Information Structure Effect on Dative-Accusative Order in Korean. *Discourse and Cognition* 15, 127–152.
- Collins, Peter. 1995. The Indirect Object Construction in English: An Informational Approach. *Linguistics* 33, 35–49.

- Dryer, Matthew. 1980. The Positional Tendencies of Sentential Noun Phrases in Universal Grammar. *The Canadian Journal of Linguistics* 25, 123–195.
- Ferreira, Victor S. and Hiromi Yoshita. 2003. Given-New Ordering Effects on the Production of Scrambled Sentences in Japanese. *Journal of Psycholinguistic Research* 32, 669–692.
- Gibson, Edward. 1998. Linguistic Complexity: Locality of Syntactic Dependencies. *Cognition* 68, 1–76.
- Gibson, Edward. 2000. The Dependency Locality Theory: A Distance-based Theory of Linguistic Complexity. In A. Marantz, Y. Miyashita, and W. O'Neil (eds.), *Image, Language, Brain*. MIT Press, Cambridge, pp. 65–126.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamins, Amsterdam.
- Givón, Talmy. 1988. The Pragmatics of Word Order: Predictability, Importance and Attention. In M. Hammond, E. A. Moravcsik, and J. Wirth (eds.), *Studies in Syntactic Typology*. John Benjamins, Amsterdam, pp. 243–284.
- Gundel, Jeanette K. 1988. Universals of Topic-Comment Structure. In M. Hammond, E. A. Moravcsik, and J. Wirth (eds.), *Studies in Syntactic Typology*. John Benjamins, Amsterdam, pp. 209–239.
- Hawkins, John. 1994. *A Performance Theory and Constituency*. Cambridge University Press.
- Hawkins, John. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press.
- Kim, Hongkyu. 2000. The 21st Sejong Project: Modern Korean Balanced Corpus. [<21 세 기 세종계획> 국어 기초자료. 연구·교육용 1,000 만 어절 현대국어 균형 말뭉치], CD-ROM.
- Rickford, John, Thomas Wasow, Norma Mendoza-Denton, and Juli Espinoza. 1995. Variation and Change in *as far as* Constructions. *Language* 71, 102–131.
- Siewierska, Anna. 1993. Syntactic Weight vs. Pragmatic Factors and Word Order Variation in Polish. *Journal of Linguistics* 29, 233–265.
- Snyder, Kieran. 2003. *The Relationship Between Form and Function in Ditransitive Constructions*. Ph.D. thesis, University of Pennsylvania.
- Temperley, David. 2007. Minimization of Dependency Length in Written English. *Cognition* 105, 300–333.
- Wasow, Thomas. 1997. Remarks on Grammatical Weight. *Language Variation and Change* 9, 81–105.
- Wasow, Thomas. 2002. *Postverbal Behavior*. CSLI Publications, Stanford.
- Wasow, Thomas. 2007. Selecting Among Paraphrases. Presented at the University of Konstanz, Germany.
- Yamashita, Hiroko and Franklin Chang. 2001. “Long before Short” Preference in the Production of a Head-Final Language. *Cognition* 81, B45–B55.
- Zec, Draga and Sharon Inkelas. 1990. Prosodically Constrained Syntax. In S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection*. CSLI Publications, Stanford, pp. 365–378.

Submitted on: April 15, 2009

Accepted on: May 29, 2009