Taking Hebrew into account

Introduction

Arnold, Wasow, Losongco and Ginstrom (2000) show that in English, two independent factors - grammatical complexity (heaviness) and discourse status (newness) - influence word order in the Heavy NP Shift and Dative Alternation constructions, and neither can be explained away using the other. The purpose of this paper is to replicate this study for a sub-case of the Heavy NP Shift construction, by concentrating on two Hebrew collocations: *lakaxat bexeSbon* ('take into account') and its synonym *lehavi bexeSbon* ('take into account', lit: 'bring into account'), and *lehavi liyedi'a.*. ('call to (someone's) attention'). Both collocations can appear in a "non-shifted" constituent order (V + NP + PP) as well as in a "shifted" order (V + PP + NP) without change in meaning, as can be seen below:

(1) a. *kxu et kol ha-macavim be-xeSbon*
take ACC all the-situations into-account
take every situation into account.
(http://www.com-line.biz/, retrieved August 1st 2009).

b. *kxu be-xeSbon et kol ha-macavim*
take into-account ACC all the-situations
take every situation into account.

(2) a. *bikaSti lehavi et tguvati liyedi'at kor'av*
I.asked to-bring ACC my.response to.attention.of his.readers
I asked to call my response to his readers' attention
(http://www.notes.co.il/joseph/24935.asp, retrieved July 31st 2009).

b. *bikaSti lehavi liyedi'at kor'av et tguvati*
I.asked to-bring to.attention.of his.readers ACC my.response
I asked to call my response to his readers' attention.

Following Wasow (1997), Arnold et al. suggest that collocations are generally more prone to shifting than non-collocations. This is traced back to speakers' tendencies to keep all parts of the collocation frozen together, rather than split them apart across the utterance. Nonetheless, Arnold et al. show that alongside this general tendency, both heaviness and newness are factors in deciding constituent order in English collocations.

In this paper, I will show that Hebrew behaves differently than English, in that heaviness alone influences the constituent order in both studied collocations, and newness does not play a significant role. Additionally, I will argue that similarly to English, Hebrew collocations have a higher tendency to shift than is predicted by heaviness alone. Following the argumentation Arnold et al. provide for English, I will suggest that other factors beside heaviness may be responsible for this tendency.

1 Explaining constituent order in Arnold et al.

Arnold et al. (2000) show that two independent factors - grammatical complexity (heaviness) and discourse status (newness) influence word order in two English constructions - Heavy NP Shift and Dative Alternation. Below are examples for both types of constructions:

(3) Heavy NP Shift (HNPS)
a. The waiter brought the wine we had ordered to the table.
b. The waiter brought to the table the wine we had ordered.
(4) Dative Alternation (DA)
a. Chris gave a bowl of Mom's traditional cranberry sauce to Terry.
b. Chris gave Terry a bowl of Mom's traditional cranberry sauce.
(Arnold et al 2000: 28).

Using two different methodologies - a corpus study and an elicitation experiment, the authors show that the constituent order in both DA and HNPS is influenced by the relative newness and heaviness of these constituents. In both cases, speakers tend to produce the relatively newer and heavier constituent later. Neither factor could be statistically explained away by the other, and both were significant.

It is of conceptual interest to note that Arnold et al. argue that the motivation for this tendency can be traced back to the speaker's needs, not to the listener's, as is commonly assumed. A speaker who experiences difficulties in production will tend to opt for the "shifted" constructions. That is, production constraints are what governs the choice of constituent order.

In the corpus study, *heaviness* is measured by Arnold et al. as the difference in length between two constituents, in terms of number of words. For *Newness*, Arnold et al. distinguish between discourse-given, inferable, and discourse-new. An NP is classified as given if its referent has been previously mentioned in the discourse. An NP whose referent has not been explicitly mentioned but could be inferred from something else that was mentioned is classified as inferable. Only NPs whose referents are truly new to the discourse are classified as new.

Arnold et al. ask themselves whether grammatical structure and discourse status both affect constituent order, or whether one factor can account for all the variation. In order to answer this question, they first conducted a corpus study, in which they extracted examples of both HNPS and DA. To investigate HNPS, they searched for examples of the (possibly discontinuous) strings *bring . . . to . . .* and *take . . . into account . . .*. In investigating DA, they chose the verb *give* as a representative.

The corpus search yielded 223 examples of *bring . . . to*, 167 examples of *take . . . into account*, and 269 examples of *give*. These were coded for constituent order: The values NONSHIFTED (V + NP + PP) and SHIFTED (V + PP + NP) were assigned for HNPS, and PREPOSITIONAL (V + NP + PP) and DOUBLE OBJECT (V + NP + NP) were assigned for DA. Heaviness was measured as the relative length of the two constituents, in terms of number of words. Newness was coded as one of three categories: given, inferable, and new.

The corpus analysis showed that both heaviness and newness are significantly correlated with constituent ordering, and that neither correlation can be reduced to the other. It showed that both DA and HNPS are influenced by the relative newness and heaviness of the two constituents. In both cases, speakers tended to produce the relatively newer and heavier constituent later.

The authors note that the 'take into account' construction shifted at a greater rate than the 'bring to' construction. They argue that something about the construction itself is causing speakers to shift more in 'take into account' compared to 'bring to'. Following Wasow (1997), the authors note that collocations have a higher rate of shifting than non-collocations. Wasow shows that semantically opaque collocations, like 'take into account', have a higher rate of shifting than other collocations, for example, 'bring to an end'. Therefore, once a speaker has decided to use an idiom like 'take into account', it is easier to produce it all together, leading to a shifted structure. The authors use this explanation to account for that fact that the 'take into account' construction shifted to a much greater extent than could be predicted by either heaviness or newness, suggesting that other factors may be at play - for example the degree to which the verb and prepositional phrase constitute a fixed expression.

Armed with these results from the corpus study, Arnold et al. use a second methodology, an elicitation experiment, to confirm their results. The details of this experiment need not concern us, since this paper only attempts to replicate the corpus study. Similarly to the corpus study, the experiment provided clear evidence that both newness and heaviness have significant effects on constituent ordering, with speakers tending to produce their relatively newer and heavier constituents later in the sentence.

2 Replicating the corpus study: the case of Hebrew

In order to facilitate comparison with the results of Arnold et al., I adopt the definitions and parameters defined in their study, as described in section 1 above. I concentrate on two Hebrew collocations, in an attempt to reproduce the HNPS results: *lakaxat bexeSbon* ('take into account') and its synonymous counterpart *lehavi bexeSbon* ('take into account'; lit: 'bring into account'), and *lehavi liyedi'a.*. ('call to (someone's) attention').

2.1 'take into account'

2.1.1 The data

Because of the lack of suitable corpora, I decided to run a customized google search. I searched only for the word *bexeSbon* ('into account'), which allowed me to find all declinations of "bring", in both constituent orders, all at once. In order to narrow down the search to a reasonable size, I chose to run it on three blogs of different registers: *http://www.tapuz.co.il/* and *http://israblog.nana10.co.il/*, which use mostly colloquial language (77 results), and *http://notes.co.il/*, whose users write in a higher register (74 results). I excluded from the search words that typically appeared in unwanted results (e.g., "bank"), and went through the rest of the results one by one.

I excluded results in which the desired expression appeared in the title or in bullet points, as these tended to be shortened versions with detailed descriptions appearing later on in the text. I also excluded any text which used the same construction in a repetitive fashion - more than twice in proximity or more than three times throughout the text. Lastly, I excluded constructions that had a sentential complement- i.e., constructions of the type V+PP+S, in which the complement of 'take into account' is a sentence ("that …."). The V+PP+S construction cannot shift in Hebrew, as can be seen in examples (5)-(6), and therefore its inclusion would skew the results.

(5) a. *tamid carix lakaxat be-xeSbon Se-lo kulam yavinu et ha-puenta*
always must take into-account that-not everybody understand ACC the-point
You must always take into account that not everybody is going to get the point
(http://www.notes.co.il/avi/, retrieved July 29th 2009).

b. **tamid carix lakaxat Se-lo kulam yavinu et ha-puenta be-xeSbon*

(6) a. *kxu be-xeSbon Se-im nihiye kvuca gdola, uxal ulay lehozil et ha-mexir*
take into-account that-if we-will-be group large I-can maybe reduce ACC the-price
Take into account that if we'll be a large group, I might be able to lower the price
(http://israblog.nana10.co.il/blog=22501, retrieved July 29th 2009).

b. **kxu Se-im nihiye kvuca gdola, uxal ulay lehozil et ha-mexir be-xeSbon.*

All in all, I collected 151 suitable examples of the 'take into account' construction. These were coded as either NONSHIFTED (V + NP + PP) or SHIFTED (V + PP + NP); They were also coded for newness and heaviness: newness was coded as one of three categories: given, inferable, and new. Heaviness was measured as the relative length of the two constituents, in terms of number of words (that is, the number of words in the NP minus 1). I excluded from my count the accusative marker *et*, which appears before definite direct objects (DOs). Only the direct object was coded for newness, since newness is not applicable to prepositional phrases like into 'into account'.

2.1.2 Results

The purpose of this paper is to replicate part of the corpus study conducted by Arnold et al.: it aims to investigate the effects of newness and heaviness on constituent ordering in Hebrew collocations and compare the result with the original results of the English study. In order to replicate the Arnold et al. study as closely as possible, I analyzed the relationship of each factor to constituent ordering using the same statistical tests employed by Arnold et al., using SPSS 11.5. As in their analysis, the statistical test performed was the logistic regression, chosen since the dependent variable is dichotomous. (see Arnold et al for further explanation of the methods).

Since I had many tokens in the *heaviness* cells, I could use my data without grouping my results together, except for the case of NPs weighing 8 words or more. These were all collapsed into a single category. My categorization differs from that of Arnold et al., as I had no results in which the PP was heavier than the NP. The categorization is shown in table 1:

a) DO = PP	DO length - PP length =	0
b) DO > PP	DO length - PP length =	1
c)	DO length - PP length =	2
d)	DO length - PP length =	3
e)	DO length - PP length =	4
f)	DO length - PP length =	5
g)	DO length - PP length =	6
h)	DO length - PP length =	7 or more

TABLE 1. Categories of heaviness for 'take into account'.

Very few tokens were coded as *inferable* (7 NPs out of 151), rendering this category too small for statistics. Therefore, I decided to group this category together with the given information, similarly to the decision made by Arnold et al.

My main finding is that unlike English, **only heaviness significantly affects constituent ordering in the Hebrew collocation 'take into account'** (-2 * Log LR = 40, $p < .001$). Arnold et al. show that in English, both heaviness and newness are significant factors in deciding constituent ordering in collocations. Nonetheless, newness did *not* have a significant value ($p < .234$) in my regression, and I therefore conclude that it does not have an effect on constituent ordering in the Hebrew collocation 'take into account'. That is, speakers tend to use the shifted order more frequently when the direct object is heavy, regardless of whether or not it is discourse-new.

Figure 1 shows the collocation's tendency to shift: it is clear that the heavier the NP, the likelier it is that the constituent order will shift. Note that whenever the direct object contains more than a single word, the collocation has at least a 94% percent chance of shifting. Whenever the direct object contains 4 words or more, the ordering always shifts.

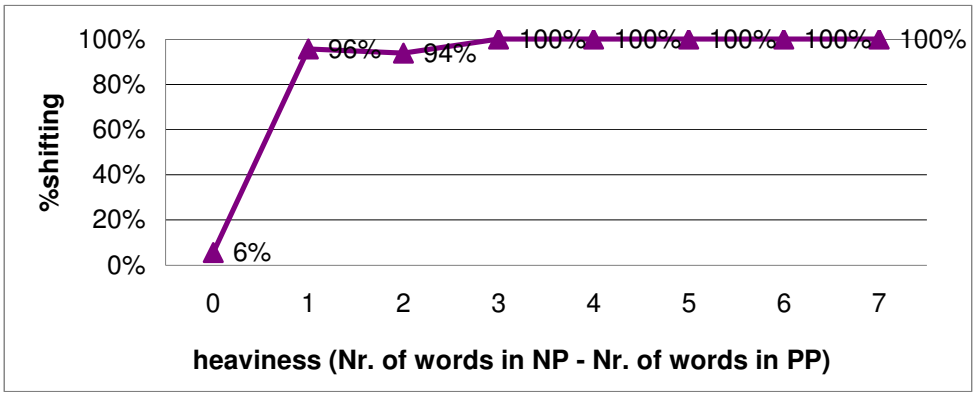

FIGURE 1: 'take into account' results

Following the observations of Wasow and Arnold et al., it is clear that in Hebrew too, 'take into account' shifts to a much greater extent than heaviness alone can explain. However, unlike in English, newness does not seem to play any significant role in influencing the constituent order in the Hebrew collocation. This can be seen in figure 2, which shows the shifting rate of discourse-new DOs compared to discourse-given DOs. In both cases, the shifting tendency is similarly maintained, with newness contributing little-to-no information as to what the order will be.

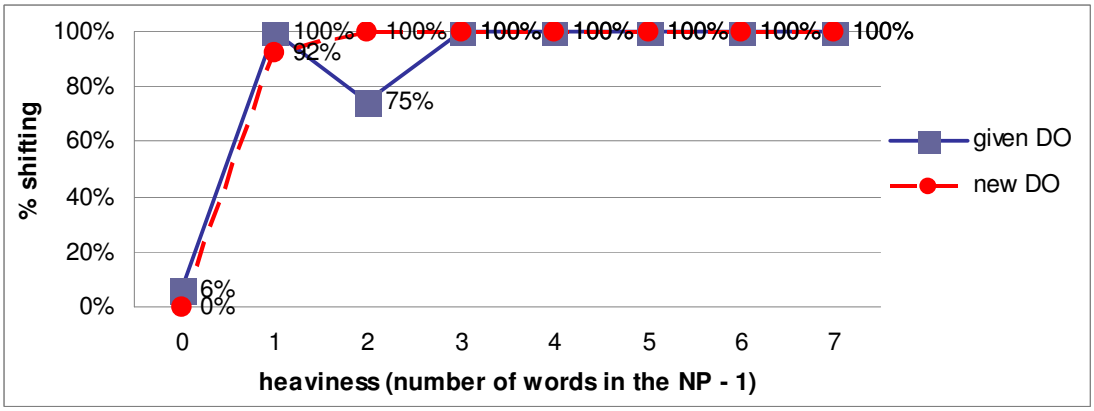

FIGURE 2: 'take into account' results broken down according to the DO's discourse status

2.2 'call to attention'

2.2.1 the data

For the second half of the corpus study, I chose to concentrate on the collocation *lehavi liyedi'a..* ('call to (someone's) attention'). My choice was based on theoretical, as well as practical, considerations: on the theoretical side, unlike the first collocation, this one has an open slot. The explicit specification of the person(s) whose attention is being called is required, but is not restricted to a predetermined set. In other words, *lehavi liyedi'a..* allows me to test for the effect of both the NP's newness and the PP's newness. This can be seen in example (7), in which, in the appropriate contexts, the various NPs and PPs which appear in the sentence can be either new or given.

(7) *ha-iton hevi et {ha-inyan, ha-sod, ma Se-roS ha-memSala nisa*
the-paper called ACC{the-issue, the-secret, what that-prime the-minister tried
lehastir, ne'umo Sel ha-sar…} liydi'at {anaSim nosafim, ha-cibur,
to-hide, the.speech of the-minister} to-the-attention-of people additional the-public,
kor'av, -xem…}
its.readers, -your…}
The paper called {the issue, the secret, what the prime minister tried to hide, the minister's speech...} to {additional people's, the public's, its readers', your...} attention.

This variation, alongside two additional factors, make 'call to attention' less semantically opaque than 'take into account': while *be-xeSbon* ('into account') is taken as a single chunk and cannot be replaced with anything in 'take into account', *li-yedia..* ('to attention') can be replaced with *le-muda'ut* ('to awareness') in 'call to attention' without change in meaning, making the collocation less fixed. In addition, *xeSbon* ('account') is not interpreted as a mathematical account, and therefore 'take into account' is not interpreted compositionally. By comparison, *yedia* ('attention') is interpreted literally, and 'call to attention' is thus compositional; Since 'call to attention' is less semantically opaque than 'take into account', there should be less pressure on it to keep all its idiomatic parts together. Hence, it should be less likely to shift, and could be expected to be more affected by newness and heaviness than the latter collocation turned out to be.

On the practical side, I chose the collocation 'call to attention' because it was relatively easy to narrow down the google search for *lehavi liyedi'a..* to a reasonably sized pool from which most of the irrelevant results could be eliminated. This was achieved by searching for the expression *liyedi'a** (the joker allowed for all continuations of the PP).

I used the same methodology as for the 'take into account' data: I conducted my search on the same sites - *http://www.tapuz.co.il/* and *http://israblog.nana10.co.il/* (57 results), and *http://notes.co.il/* (65 results), so as to sample the same database for both parts of my study. Here too, I excluded from the search words that typically appeared in irrelevant results, and went through the rest of the results one by one. I excluded results in the same manner and for the same reasons as described in section 2.1.1.

All in all, I collected 123 relevant examples of the *lehavi liyedi'a..* construction. I coded these as either NONSHIFTED (V + NP + PP) or SHIFTED (V + PP + NP); I also coded them for newness and heaviness using the same categories as before: Heaviness was measured as the relative length of the two constituents (that is, the number of

words in the NP minus the number of words following *liyedi'a* ('to attention') in the PP). Newness was coded as either given, inferable, or new. This time, both the NP and the PP were coded for newness. I decided to code expressions referring to public figures (e.g., "the mayor"), public institutions (e.g., "the police"), collective terms (e.g, "the public", "everyone") and private names, as given information. This decision is only relevant for the 'call to attention' data, as the 'take into account' data did not contain similar tokens.

2.1.2 Results

Because of the small number of tokens in some cells, I had to collapse my results and decided to use the categorization of heaviness defined by Arnold et al. :

a) DO << PP	DO length - PP length =	-4 or less
b) DO < PP	DO length - PP length =	between -1 and -3
c) DO = PP	DO length - PP length =	0
d) DO > PP	DO length - PP length =	between 1 and 3
e) DO >> PP	DO length - PP length =	4 or more

TABLE 2. Categories of heaviness for 'call to attention'.

Very few tokens were coded as *inferable* (2 NPs out of 246), rendering this category too small for statistics. Therefore, I decided to group it with the given information, as I did for the 'take into account' data above.

My main finding here as well, is that unlike English, **only heaviness significantly affects constituent ordering in the Hebrew collocation 'call to attention'** (-2 * Log LR = 117, $p < .001$). As before, though Arnold et al. conclude that both heaviness and newness are significant factors in determining constituent ordering in English collocations, in my regression, neither the NP's newness nor the PP's newness had significant values ($p < .08$ and $p < .035$, respectively). Therefore, I conclude that newness does not have an effect on constituent order in the Hebrew collocation 'call to attention'. That is, speakers tend to use the shifted order more frequently when the direct object is heavy, regardless of whether or not it is discourse-new.

Figure 3 shows the results for the 'call to attention' data. The shifted constituent order is more frequent when the direct object is relatively longer than the PP. Note that whenever the direct object is at least 4 words longer than the PP, there is a very great chance (97%) that the constituent order will shift.

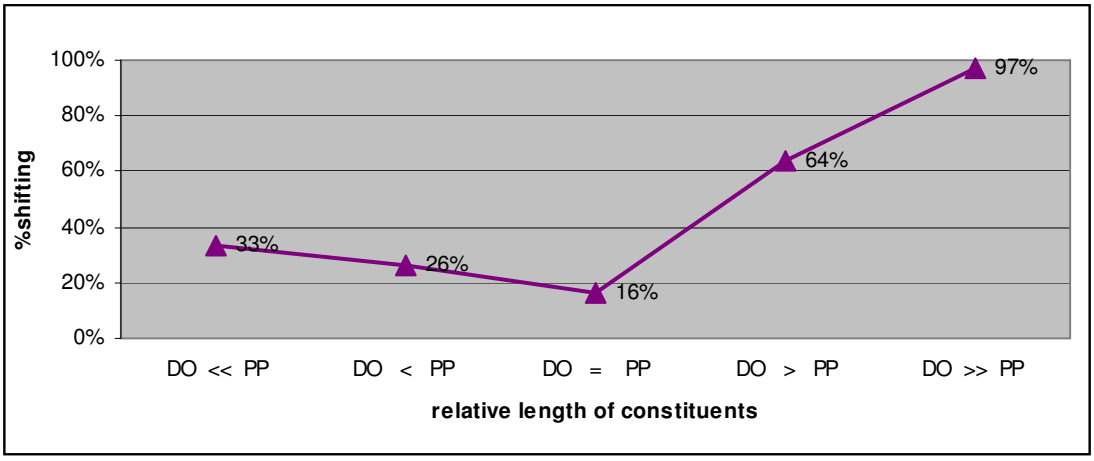

FIGURE 3: 'call to attention' corpus analysis results

3 Discussion

The results of the study show that Hebrew resembles English in some respects, but differs in others: while both heaviness and newness significantly influence constituent ordering in English collocations (and in HNPS constructions as a whole), only the former, but not the latter, influences constituent order in Hebrew collocations. That is, speakers tend to produce the relatively heavier constituent later, regardless of whether it is discourse-new or discourse-given.

Like the case of English, in Hebrew too, heaviness (and newness) alone cannot explain the large amount of cases in which speakers produce a shifted construction. It seems that other factors are involved, which increase the probability that the collocations will shift. In both languages, the more idiomatic collocation, 'take into account', is likelier to shift than the other collocation tested ('bring to' for English, and 'call to attention' for Hebrew). In this too, Hebrew behaves similarly to English.

I believe that Wasow's explanation for English neatly extends to Hebrew as well. He argues that since speakers tend to keep all parts of the frozen collocation together, collocations are generally more prone to shifting than non-collocations. Furthermore, the less semantically transparent a collocation is, the likelier it is to shift. Once a speaker decides to use a relatively opaque idiom like 'take into account', it is easier for her to produce it all together, leading to a shifted structure. In Hebrew, 'take into account' is more semantically opaque than 'call to attention' for the reasons specified in section 2.2.1 - 'take into account' is fixed and non-compositional, while 'call to attention' (a) is interpreted compositionally; (b) is less fixed, since *li-yedia*.. ('to attention') can be replaced with *le-muda'ut* ('to awareness') without change in meaning; and (c) has an extra empty slot that for specifying the person(s) whose attention is being called, which is not restricted to a predetermined set. As Wasow predicts, 'take into account' indeed shifts at a greater rate than 'call to attention' (77% of the corpus in the former case compared to 56% in the latter). Therefore, one might consider the degree of semantic transparency of a collocation as an additional factor that affects the ordering of its constituents, beside heaviness and newness.

In thinking of other factors that might affect constituent ordering in collocations, it might be useful to consider Bresnan's recent work on dative alternation in American, Australian and New Zealand English. In a series of papers, Bresnan and various colleagues (Bresnan 2007, Bresnan & Nikitina 2007, and Bresnan, Cueni, Nikitina & Baayen 2007) have argued that the double object construction (8a) and the prepositional dative construction (8b) are essentially the same, and there is no structural/semantic difference between them. Speakers choose which one to use based on a number of factors, such as definiteness, animacy, pronominality, discourse accessibility, heaviness, newness, and so on. They conclude that speakers' syntactic knowledge is probabilistic, and that these factors and probabilities are part of the grammar.

(8) a. We give children candy.
    b. We give candy to children.

Bresnan et al. use a probabilistic model of corpus dative productions and a set of psycholinguistic experiments to measure the rate at which American and Australian subjects choose to produce both structures. They find that the Australian subjects

share with the American subjects a sensitivity to corpus probabilities, but that they differ in how they rate the different variables that affect their selection.

It is outside the scope of my current study to test which of the variables mentioned above influence constituent order in Hebrew collocations, but it seems very likely that at least some of them do. Examining my corpus, one variable that is clearly strongly correlated with constituent order is pronominality (i,e., whether or not the direct object is a pronoun). In 91% of the cases in which 'take into account' appeared in a non-shifted order, the direct object was a pronoun. By comparison, less than 1% of the shifted constructions contained a pronoun. Since pronouns are also very short, this might suggest that heaviness is in fact not a factor that independently influences constituent order, or at least not as strongly as my study currently predicts, but rather that it is mediated by pronominality: whenever something is a pronoun, it is likely to appear in the non-shifted construction. It just happens that pronouns are also short, thereby making heaviness seem as though it influences the choice of order.

While this finding weakens the conclusion that heaviness alone influences constituent order in Hebrew collocations (or at least its extent), it strengthens the conclusion that newness is *not* a significant factor in the choice. In 94% of the cases in which 'take into account' appeared in the non-shifted order, the direct object contained given information. Nonetheless, despite this seemingly strong correlation between newness and the non-shifted construction, newness was determined to be an insignificant factor. Note that in 91% of the cases in which 'take into account' appeared in the non shifted construction, the direct object was a pronoun (and thus given information). Since pronominality is likely to cancel out at least part of the effect of newness, the conclusion that it doesn't affect constituent order is strengthened.

In examining 'call to attention', pronominality seems to have less of en effect. Only in 39% of the cases in which 'call to attention' appeared in the non-shifted order was the direct object a pronoun. Additionally, more than 11% of the shifted constructions contained a pronoun. Over 90% of the non-shifted constructions contained given information, but in only 43% of those cases was the direct object a pronoun. Therefore, it seems clear that other factors are at play here, which remain to be discovered in future work.

The finding that other factors beside heaviness, such as pronominality and degree of transparency, are at play in deciding the constituent order in Hebrew collocations, is an interesting one. It might serve to strengthen the Bresnan et al. conclusion from the comparative study of American and Australian English: different languages are sensitive to the same factors, but differ in how they rate their relative weight. Thus, in deciding constituent ordering in collocations, Hebrew might rate newness low enough to render it insignificant as an independent influencing factor, because its interaction with other, higher ranked factors, mediates most of its weight in the decision. By comparison, American English rates newness higher, rendering it a significant factor in the decision.

Conclusion

The main conclusion of this paper is that constituent order in Hebrew collocations is influenced by partially different factors than English collocations. I've come to this conclusion after replicating the corpus study conducted by Arnold et al. on a Hebrew corpus that I created. I concentrated only on part of the phenomena that Arnold et al. tested, namely on collocations, which are a sub-case of the Heavy NP Shift construction. To the extent that it was possible, I used the same methodology and tests used by Arnold et al., but came to a partially different conclusion from theirs.

While Arnold et al. find that both heaviness and newness influence constituent ordering in English collocations, I found that in Hebrew, only the former, but not the latter, has a significant effect. Similarly to English, it is clear that the collocations shift to a much greater extent than heaviness (and newness) alone can predict, thus making the case for claims such as that of Bresnan et al., that many other factors are at play here. Examining my data, I make the case that pronominality and the degree of transparency of the collocation could be two such factors, and that other factors must be found. However, finding these factors is outside the scope of this paper, and I thus leave it for future research. Lastly, I suggest that my results, combined with future research, might lend strength to the conclusion of Bresnan et al. that languages share the factors that affect decisions such as constituent ordering in collocations, but differ in how they rate their relative weight.

References

Arnold, J., Wasow, T., Losongco, A., Ginstrom, R., 2000. Heaviness vs. newness: the effects of complexity and information structure on constituent ordering. *Language* 76, 28–55.

Bresnan, J., 2007. Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation. In S. Featherston and W. Sternefeld, eds., *Roots: Linguistics in Search of its Evidential Base*, Berlin: Mouton de Gruyter, pp. 75–96.

Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen, 2007, Predicting the Dative Alternation. In G. Boume, I. Krämer and J. Zwarts, eds., *Cognitive Foundations of Interpretation*, Amsterdam: Royal Netherlands Academy of Science, pp. 69–94.

Bresnan J., and T., Nikitina. 2007. The Gradience of the Dative Alternation. In L. Uyechi and L. Hee Wee, eds., *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. Stanford: CSLI Publications.

Wasow, T., 1997. Remarks on grammatical weight. *Language Variation and Change* 9, 81–105.

Hadas Kotek
Tel-Aviv University, August 2009.