



GÖTEBORGS UNIVERSITET
INST FÖR SPRÅK OCH LITTERATURER

ENGLISH

Watch your weight!

**A corpus-based study of end-weight constructions in the speech of
native speakers and Swedish learners of English**

Author: Elias Ottmar

Course: EN1321 Linguistic Research Project, 15 HEC

Date: 15/6 2015

Supervisor: Viktoria Börjesson

Examiner: Larisa Gustafsson Oldireva

Department: University of Gothenburg, Dept. of Languages and Literature, English

Course: EN1321 - English: In-depth course in English, Linguistic Research Project, Bachelor's Degree Essay Project (C-level paper)

Title: Watch your weight! A corpus-based study of end-weight constructions in the speech of native speakers and Swedish learners of English

Author: Elias Ottmar

Supervisor: Viktoria Börjesson

Abstract

When we engage in conversation, it is necessary to order our speech so that others can understand us. To avoid potentially problematic internal elements, there is a tendency to place heavy clause constituents in sentence-final position. This is known as end-weight. This study investigates how spoken language clause constituent order functions for two weight-sensitive constructions - heavy noun phrase shift and dative alternation - in Swedish learner and native English speaker university students. Which factors determine the position of the constituents and whether there are differences between the native and learner speakers is also investigated. The data used is extracted from the Swedish component of the LINDSEI corpus and its native counterpart, LOCNEC. Chosen two- and three-placed verbs were located in the data to find weight-sensitive constructions and the results analysed quantitatively and qualitatively. The results show that end-weight shifting is noticeably resistant to syntactic factors, such as complexity and length, which have little influence on clause constituent ordering in both the native and the learner speakers. Semantic and pragmatic motivators are discussed as having potential influence on word order: hesitation markers can indicate disruptions in speech production and lexical bias can influence speakers' choice of structure. The facilitation of language acquisition between similar languages can possibly provide an explanation for the lack of differences between the learner and the native data. The essay concludes that many factors can have potential influence over weight-sensitive constituent order.

Keywords: end-weight, information structure, dative alternation, heavy NP shift, corpus study, foreign language

Contents

1. Introduction	4
1.1 General introduction	4
1.2 Research aim and questions	4
1.3 The structure of the essay	5
2. Theoretical background	6
2.1 Defining end-weight	6
2.1.1 Criteria	6
2.1.2 Complexity and length	7
2.2 End-weight phenomena	7
2.2.1 Heavy noun phrase shift	7
2.2.2 Dative alternation	8
2.3 Semantic and pragmatic approaches to end-weight	8
2.3.1 Speaking-oriented vs. listener-oriented approaches	8
2.3.2 Information structure and language acquisition	9
3. Previous studies	11
3.1 Syntactic motivators for end-weight shifting	11
3.2 Semantic and pragmatic motivators for end-weight shifting.....	12
4. Methodology	14
4.1 Material: the corpus	14
4.1.1 Corpus studies	14
4.1.2 LINDSEI-SW and LOCNEC	15
4.2 Method	15
5. Results	17
5.1 Heavy noun phrase shift	17
5.2 Dative alternation	20
6. Analysis and discussion	23
7. Conclusions and further research	27
References	28

1. Introduction

1.1 General introduction

When we engage in conversation, it is necessary to arrange information in our speech so that others can follow us. One feature that influences the structure of information is the use of *end-weight*: that is, a speaker's tendency to order clause constituents so that large and heavy elements are positioned toward the end of a sentence (Wasow 1997a). These parts usually contain information that is new in the discourse context, since there is a preference for receiving new information as late as possible (Arnold and Lao 2008). This tendency is hardly disputed, but the clause constituent order is not only about making yourself understood on a grammatical level, but also about facilitating the communicational exchange between speaker and listener in terms of, for example, how we refer to entities in the discourse.

To a native English speaker, English end-weight structure tends to come naturally, but to a person using English as a foreign language, it might pose more of a problem if the native language differs from English in terms of how clause constituents are typically ordered (Hawkins 1990; Siewierska 1993). For example, little research has been done on how Swedes handle information structure when speaking English, and nothing has been done on end-weight. This study therefore aims to fill that gap by looking at how clause constituent order is constructed in terms of end-weight in Swedish learners' speech. The results can indicate the extent (or, possibly, lack) of Swedish familiarity with aspects of English syntax that are perhaps not necessary for being understood, but could affect the impression made on listeners or readers in terms of how fluent, or natural, or "smooth", the speaker's use of English is.

1.2 Aim and research questions

The aim of the study is to investigate how Swedes who speak English as a foreign language construct their sentences in relation to the end-weight principle. The following research questions have been formulated, whose answers ought to be helpful in reaching the aim:

- How are different types of constituents (noun phrases, prepositional phrases etc.) ordered?
- Which factors determine the relative positions of the constituents?

- Does Swedish learners' use of English clause constituent order differs from that of native English speakers, according to the principle of end-weight? If it does, what are the differences?

1.3 The structure of the essay

Firstly, relevant aspects and perspectives are outlined in the theoretical section. Different varieties and functions of end-weight are explained. In the methodology chapter, the material used in the study is presented and the method with which the study has been conducted is accounted for. After that, the results are presented and discussed in the light of the given theoretical framework. Lastly, a summary with conclusions is given along with suggestions for future research.

2. Theoretical background

In this chapter, the theoretical background for the study will be outlined. Relevant terms and concepts regarding end-weight will be explained, as well as approaches that may provide an explanation for why end-weight occurs. Lastly, some concerns regarding the universality of end-weight will be addressed.

2.1 Defining end-weight

2.1.1 Criteria

Contrary to what one might think, the English language allows for considerable variation in the ordering of clause constituents where outcomes still provide the same semantic interpretation (Wasow and Arnold 2003). Such variation is especially prominent in post-verbal constituent ordering. Namely, there is a tendency to order such constituents after *weight*, or “heaviness” (Wasow 1997a: 81). Otto Behagel (1972) described this phenomenon already in 1909, stating that clauses with heavier grammatical weight tend to occur after clauses with lighter weight (via Wasow 1997a: 82). Quirk et al. coined the term *end-weight* to describe “the tendency to reserve the final position for the more complex parts of a clause or sentence” (Quirk et al. 1972: 943).

Several suggestions on what signifies weight in constituents have been proposed, their nature ranging from phonological to syntactic and semantic. We can categorise such ‘criteria’ as absolute or relative, meaning those that measure the characteristics of a single constituent and those that compare one constituent with others in the same sentence. None of these suggestions, however, are able to account for all instances of end-weight; indeed, shifted sentences occasionally occur when none of the criteria seem to apply (Wasow 1997a).

As this indicates, it is difficult to propose a water-tight definition of end-weight and to pin down exactly what factors are at work to trigger a shifted order. Such a definition is not the purpose of this essay, either. However, several studies have shown that both relative and absolute weights are relevant factors when speaking of end-weight, and thus a brief presentation of these terms is necessary for understanding the results of this study.

2.1.2 Complexity and length

There are two ways in which a constituent's grammatical weight can be syntactically measured: the first factor is *length*, as in the number of words, and the other one is *complexity*. Both length and complexity can be measured in an absolute (or categorical) sense, where only the characteristics of any single constituent are taken into account. For length, this is very straight-forward: we simply count the number of words that the constituent contains. In terms of complexity, a constituent is analysed as a simple noun (or prepositional) phrase, or as containing an embedded phrase or other modification (Wasow 1997a). A constituent that contains such elements is considered complex, while a non-modified noun- or prepositional phrase is considered simple. Weight can also be measured in a relative (or gradient) sense, where constituents are compared with one another to find which one is the longest or most complex (Hawkins 1994). Naturally, length and complexity do not work independently of one another but may both influence constituent ordering to various extents (Stallings et al. 1998).

2.2 End-weight phenomena

There are several circumstances under which end-weighted shifting can occur. In this study, two of the most well-studied variations, *heavy noun phrase shift* and *dative alternation* will be taken into consideration. For each phenomenon, there is a basic (unshifted) order of constituents as well as a shifted order that occurs when a heavy constituent is present. Because of this, such sentence-constructions are said to be weight-sensitive (Wasow 1997a).

2.2.1 Heavy noun phrase shift

Heavy noun phrase shift (HNPS) occurs when a bulky direct object noun phrase (NP) is moved to the end of a sentence, thereby switching places with a lighter or shorter constituent which is usually a prepositional phrase (Wasow 1997b; Stallings et al. 1998). An example:

- (a) I would like to present | a complete set of our latest production pamphlets | to you
- (b) I would like to present | to you | a complete set of our latest production pamphlets

In (a), the direct object NP *a complete set of our latest production pamphlets* precedes the prepositional phrase *to you*, while in (b), it is instead placed in final position (shifted order).

HNPS seems to be more preferred in those instances when the shifted constituent has a complex structure, for example an embedded clause, than when it has a simpler structure

(Wasow and Arnold 2003). When the object NP is short, however, an unshifted order is preferred (Stallings et al. 1998).

2.2.2 Dative alternation

Dative alternation (DA) allows for two different constructions for certain three-place verbs with ditransitive use: the double object construction (*I gave the boy some chocolate*) or prepositional object construction (*I gave some chocolate to the boy*) (Hawkins 1994). In the first order¹, the indirect object NP precedes the direct object NP, whereas in the other construction, the direct object precedes a prepositional phrase which includes a prepositional object (in this case, “the boy”).

Alternation is not possible in all instances, however. Verbs that signify processes of movement toward a goal prefer the prepositional construction, while processes that signal a change of state within the possessor take the double object construction (Bresnan et al. 2007: 4). In the example with the chocolate above, alternation is free because the meaning can be interpreted as conveying either meaning (the transfer of possession of chocolate from the speaker to the boy or the goal-oriented movement of handing the chocolate over), but a construction such as **I reported the police the robbery* (Hawkins 1994: 212-213) is unacceptable because there is no change of state within the recipient. Since some verbs favour one construction over the other, lexical bias obviously has an influence on which orders are available. Even so, bias can be “overridden” by the end-weight principle, resulting in authentic sentences like *give the creeps to people who hate spiders* (Bresnan et al. 2007: 7²).

2.3 Semantic and pragmatic approaches to end-weight

2.3.1 Listener- vs. speaker-oriented approaches

When considering theories on the underlying motivation for clause constituent ordering, it is useful to distinguish between speaker-oriented and listener-oriented approaches. Listener-oriented theories dictate that heavy clause elements are preferably put in sentence-final position and avoided in mid-position, as the latter risks confusing the listener by introducing potentially distracting sideways in the conversation (Stallings et al. 1998).

¹ Hawkins (1994) considers the prepositional order to be the basic order, which is converted to the double object order by the “Dative-Accusative Shift” (212).

² The unmarked order would be double object construction, but such an order would be awkward in leaving a complex NP (*people who hate spiders*) hanging in the middle. The sentence was found by this author on the online website <http://www.truthorfiction.com/cactus/> (accessed on 22th May 2015).

There has been a previous main focus on performance theories that analyse a linear string of words piecemeal, with the facilitation of the listener's comprehension in mind. According to the Early Immediate Constituent (EIC) recognition principle, listeners prefer that word order which allows them to identify the various syntactic elements of the sentence as quickly as possible (Hawkins 1990; Siewierska 1993). Regarding end-weight, this means that all the words that introduce a new element, like a phrase, ought to come as early as possible. Doing so results in a dismissal of orders with heavy internal constituents and favours end-weighted constructions (Hawkins 1994: 214).

As a contrast, speaker-oriented approaches assume that the underlying motivator for word order is, at least mainly, the facilitation of speech production and planning time of the forthcoming part of the sentence (Wasow 1997b). If we take as an example *I gave to Mary the big, valuable book by that new writer*, inserting the prepositional phrase (PP) *to Mary* before the noun phrase (NP) *the big, valuable book by that new writer* would buy the speaker some extra time compared to *I gave the big, valuable book by that new writer to Mary*, where the recognition of the PP is delayed due to the length of the intervening heavy NP.

2.3.2 Information structure and language acquisition

Despite the difficulties associated with providing a clear explanation for end-weight, accounting for the acceptability of end-weighted sentences must still be considered fairly intuitively clear for most native speakers (Wasow 1997b). The notion of universal grammar entails that "certain aspects of language must be innately present in the first language learner" (White 1990: 124), aspects usually referred to as universals, or universal tendencies. Suggestions have been made that structure-dependence is a general universal principle rather than language specific (Hawkins 1990), and that word order falls within the cognitive realm. As a reversed end-weight tendency of 'long first, short last' has been observed in head-final languages, right-branching ('short first, long last') end-weight is a tendency for head-initial languages (Hawkins 1990), of which both English and Swedish are members.

It has been suggested that end-weight is one of several functions that can be derived from universal tendencies, partly because of the "paucity of evidence concerning their properties and restrictions in the course of acquisition" (Rochemont and Culicover 1990: 149). Research has had different starting points depending on whether second language learners are considered as still having access to universal tendencies in adult age, or if they construct new models based on their native language (White 1990). Second language acquisition is used to account for the stages through which the learner acquires knowledge about the target

language, and can be facilitated by similarities in the structural characteristics of the native and target languages (White 1990).

Clause constituent ordering can be put in a larger context that treats information structure, which is also thought to be if near-enough universal, or at least very widespread (Bock and Irwin 1980). In information structure theory, a crucial general principle is that given information ought to be positioned before new, or “the tendency to place new information towards the end of the clause” (Quirk et al. 1972: 943). In other words, information availability in sentence production has syntactic effects (Bock and Irwin 1980). For example, “in the double object construction, given constituents precede constituents that are new” (Krifka 2008: 39). Generally, “the heavy-NP-shifted word order decreases the expectancy of something given as the referent of the direct object noun phrase, and increases the expectancy of something new” (Arnold et al. 2004: 287). As focus and reference change with the context, such approaches to weight-sensitive phenomena consider them to be mainly pragmatic in nature (Hawkins 1994). Nevertheless, as it operates in a syntactic context, one must turn to clause constituent ordering to find it.

3. Previous studies

3.1 Syntactic motivators for end-weight shifting

The existence of end-weight has been documented numerous times in various forms and shapes (Hawkins 1994; Wasow 1997a; Wasow 1997b; Rochemont and Culicover 1990; Stallings et al. 1998) and is hardly in doubt, but the reason for its occurrence is less agreed upon. In the search for motivating factors for weight-sensitive shifting, many, if not most, studies have traditionally concentrated mainly on absolute length (Kimball 1973; Stallings et al. 1998) or complexity (Ross 1967), although other factors have been studied and may interact with them, for example regional variations (Bresnan and Ford 2010) or animacy (Stallings et al. 1998). In 1975, Chomsky noted that: "... it is apparently not the length in words of the object that determines the naturalness of the transformation, but, rather, in some sense, its complexity" (Chomsky (1975:477), claiming that noun phrases containing embedded clauses are more complex than noun phrases without such clauses.

According to another study (Wasow and Arnold 2003), complexity also caused end-weight shifting predominantly when found in the direct object phrase only. The difference was greater when the constituents were considered more complex than when they were simpler. This implies that the complexity of a constituent may determine its position regardless of its weight.

Some studies have found that the relative length between the constituents seems to be more important for shifting than absolute length in any single one. For example, Hawkins (1994) discovered that the object NP needed to be significantly longer (at least three words) than the intervening element for an HNPS to occur. It was also found that in double object constructions following ditransitive verbs [V NP_i NP_j], the indirect object NP_i tended to be smaller than the direct object NP_j in number of words. The result was explained with that a double object construction "shortens the recognition domain for VP" (Hawkins 1994: 214) compared with the prepositional order, since all elements are recognised by the listener more quickly in fewer words.

Later studies question the accuracy in ascribing too much importance to any single factor for determining clause constituent order. As Shih and Grafmiller (2011) put it, "'weight' effects cannot be reduced to a single dimension". While studying dative constructions, they

found that “the number of words *can* act as a sufficient proxy for syntactic complexity and ‘weight’” (ibid. original italics), but they considered complexity to be “the *most* salient manifestation of ‘weight’” (ibid., my emphasis).

3.2 Semantic and pragmatic motivators for end-weight shifting

Speaker- and listener-oriented approaches are not necessarily in conflict with one another in practice, as in many situations both speaker and listener agree on how clause elements should be constructed in order to form a cohesive utterance, but opinions differ on which perspective should take the front seat (Wasow 1997a). In many cases, it might also be difficult, or even impossible, to measure which perspective is more likely to be at work (Stallings et al. 1998).

Hawkins’s findings on the importance of relative length (discussed in section 3.1) raise some interesting questions regarding sentence production. If speakers base their clause construction order on the relative length or complexity of all the elements involved, they should have mental access to the layout of the construction quite early on. Following an experiment where the subjects showed slightly longer decision times when uttering shifted sentences than unshifted ones, Stallings et al. suggested that the differences “could be diagnostic of competition between alternative phrase orders” (Stallings et al. 1998: 399), meaning that both orders are at least initially available for the speaker. In addition, speech disfluencies indicated that “shifting is, at least in part, a strategy invoked when the production process is particularly difficult or the shifted item is less accessible” (Stallings et al. 1998: 393).

Information that is new in the discourse context tends to be described in more detail than given information; new entities tend to be represented by larger or more complex constituents than old entities (Bock and Irwin 1990). For example, if *a tall woman with a green hat* is introduced in a conversation, she is more likely to be described using the entire NP, but if she is already established in the discourse, the full NP is more often than not superfluous (as it can be replaced, for example, by *her*). However, Wasow (1997a) has found examples of end-weighted sentences where the shifted NP cannot reasonably be considered new information in the discourse context, so it is obvious that the correlation is not direct.

There are studies whose results challenge listener-based theories for clause constituent ordering. For example, a study by Wasow’s (1997a) explored different theories from the viewpoint of how well they predict end-weight constructions. Examples were found where a [V PP NP] clause contained a verb-prepositional collocation where the semantic context was

not obvious, a so called opaque collocation. From a speaker-oriented point of view, Wasow argued, placing the PP before the NP is beneficial for both speaker and listener, even if the PP is heavier. This would be because the speaker benefits from extended speech planning time for the following NP, and the listener benefits because the meaning of the opaque collocation is only recognised when the verb and the preposition are adjacent. When the PP is shorter than the NP, however, it needs not stand before the NP from the listener's perspective, but it still can. Wasow (1997b: 352) uses the example “[t]hat will bring to the plate Barry Bonds”. This order is not beneficial from the listener's point of view; yet, it still occurs, which Wasow argues supports production-based explanations for end-weight shifting. He admits, however, that this is only “one crucial test” in deciding between the perspectives (353), and is not a universal solution.

Lexical bias has also been considered to have a significant influence on the distribution of dative alternation and HNPS (Wasow and Arnold 2003). Stallings et al. (1998) found that HNPS was less likely to occur in sentences with verbs that were more closely connected with their complements, while verbs that allowed interruption between verb and complement were more prone to allowing such an order. However, other studies have demonstrated that dative alternation “variants spontaneously occur as partial repetitions in discourse” (Bresnan and Ford 2010: 170) in the same speaker, suggesting that too much importance should not be ascribed to bias out of hand.

4. Methodology

4.1 Material

4.1.1 *Corpus studies*

A corpus is a body, or collection, of texts that have been selected with the purpose of being representative for a certain type of language. The main advantage of a corpus is that it consists of “naturally occurring language” that is used in a context and can be considered authentic, as opposed to for example constructed word lists (Gries and Newman 2013: 258). The corpus can contain written or spoken (transcribed) language from various genres, registers or styles. The main advantage of spoken language is its higher degree of naturalness compared to written language. Speakers must undergo a constant process of organising the information so that the listener will understand them, but they have less time to do so than when writing, since our utterances are spontaneous to a much greater extent in free speech. The order of the clause constituents in such material will therefore indicate how the speech is organised and thereby how the information is structured.

This study takes only spoken language material into consideration. The spoken material consists of free speech centred on certain topics, and the spontaneous nature of speech makes any corrections concerning information and clause constituent order obvious. With written materials, a text is a finished product where the stages of correction are not visible, making the process behind choosing a word order unavailable to the researcher. While written texts tend to be more formal in tone and therefore may contain larger elements that are more likely to shift, a person’s spontaneous intuition (or lack thereof) for sentence structures cannot be measured to the same extent (see chapter 6: Analysis and discussion). In a spoken language corpus, the researcher also has access to non-verbal transcribed features, which facilitate the analysis. Any corrections made by the speaker (or listener, for example for clarification) are likely to be visible in the text. Such corrections may be very valuable when considering how effective the communicative exchange is.

4.1.2 *LINDSEI-SW and LOCNEC*

This study uses two corpora: the Swedish component of *Louvain International Database of Spoken English Interlanguage* (LINDSEI-SW) and *Louvain Corpus of Native English*

Conversation (LOCNEC). The LINDSEI corpus is composed of spoken informal interviews with university students of eleven nationalities (only the interviews with the Swedish informants are used for this study, since the study treats the Swedish use of end-weight structures). The Swedish interviews are 50 in number, comprising 71,804 words (counting only the utterances of the respondents) with an average of 1,436 words per interview (Gilquin et al. 2010: 25). They were recorded between 1999 and 2005. The contents of the interviews consist of spontaneous speech around given topics (a favourite film or book, a visit to a foreign country, a specific picture sequence) as well as a free discussion, and the interviews are recorded in the presence of an interviewer who is a native speaker. The transcriptions are marked for pauses, fillers and discourse markers, but not prosodic features (Börjesson 2014: 59-60).

LOCNEC is made up of 50 informal, transcribed interviews on the same topics as in LINDSEI, but the informants are British native speakers (De Cock 2004: 227). LOCNEC was created to function as a native counterpart for the LINDSEI components and to work with the same variables so as to enable comparison between native and learner English (Hiligsmann 2014). The corpus is comprised of 117,417 words (De Cock 2004: 226), meaning that it is significantly larger than the learner data, and this difference must be taken into account when analysing statistical results. Both corpora are anonymised; all names or personal information referring to the respondents have been removed.

4.2 Method

This study makes use of both quantitative and qualitative methods, a so called mixed method approach, although with a heavier focus on qualitative analysis. Both the quantitative and qualitative concepts encompass the method for collecting the data and how the data is analysed. Quantitative analysis works with statistics to measure extent or number, while qualitative analysis is based on certain characteristics or patterns of the data that are being processed in their context (Dörnyei 2007; Rasinger 2010). Quantitative and qualitative methods are “overstated binaries” (Duff 2006: 66); there need not necessarily be any opposition between them. Indeed, in recent years, mixed method approaches have become more and more common (Rasinger 2010). In this study, the quantitative results are used as a ground for a qualitative analysis, where interesting or especially salient cases are examined more closely in terms of their characteristics.

From both corpora, data was extracted by locating verbs that are prone to occur together

with end-weight sensitive structures. In the search for heavy noun phrase shifts, 15 verbs (*add, bring, build, call, carry, draw, find, hold, keep, leave, make, put, set, take* and *write*) were chosen because they allow a transitive structure of V NP PP that could alternatively be shifted into V PP NP. For dative alternation, ditransitive uses of 3 verbs (*buy, give* and *sell*) were categorised as participating in either a double object construction with the indirect object NP preceding the direct object NP (*I gave the woman my keys*) or a prepositional construction with the direct object NP preceding an indirect object PP (*I gave my keys to the woman*). When analysing the dative alternation results, lexical bias was also looked at. The results from the native and the Swedish learner data were then compared with regard to differences or similarities between them. Due to the fairly small size of the data, as is often the case in spoken corpora research, the tokens were counted in raw frequencies.

Collocations are “grammatical patterns” (Teubert & Čermáková 2007: 23-24) of some words that commonly occur together. For this study, collocations of verbs and prepositional phrases (V-PP sequences) were interesting, for example *bring ... to, take ... into* and so on. Concordances show the relevant word in its immediate context in a specific entry in the corpus. While they give more information than collocations, since more of the context is given, the greater amount of information requires more time-consuming analysis (Gries and Newman 2008). For this reason, they are best used in studies involving a lesser amount of data, and since the corpora used in this study are small, concordances may prove helpful in some cases.

In order to conduct this study, several restrictions had to be placed on which data was used. Only those end-weight constructions mentioned above, HNPS and DA, were considered, as the time frame for this study limited the amount of corpus work that could practically be done. All lemma forms of the verbs were used in the searches. Those cases where single pronouns served as direct object for HNPS were excluded. Care was taken to separate instances where *to* functioned as a preposition from where it acted as an infinitive marker, as infinitive forms are not relevant for the study. Likewise, all utterances from the interviewer were ignored, as the focus of the study is the native and learner respondents.

5. Results

The abbreviation and number after the extracts refer to the corpus and line number where the token was found.

5.1 HNPS

Below, all found instances of the chosen verbs in transitive [V NP PP] structures are presented in Table 1. For each verb, the tokens have been counted for each corpus respectively, with the learner corpus on the left and the native corpus on the right. In Table 2, the distribution of unshifted [V NP PP] structures on collocations and non-collocations are given.

Table 1: Frequency of the verbs in [V NP PP] structure (raw frequencies)

Swe	tokens	Nat	tokens
<i>take</i>	25	<i>take</i>	27
<i>make</i>	15	<i>put</i>	17
<i>put</i>	6	<i>make</i>	9
<i>write</i>	5	<i>find</i>	8
<i>call</i>	3	<i>build</i>	5
<i>leave</i>	3	<i>bring</i>	4
<i>add</i>	2	<i>keep</i>	3
<i>build</i>	2	<i>write</i>	3
<i>find</i>	2	<i>draw</i>	2
<i>keep</i>	2	<i>hold</i>	2
<i>bring</i>	1	<i>leave</i>	2
<i>hold</i>	1	<i>set</i>	2
<i>carry</i>	-	<i>add</i>	1
<i>draw</i>	-	<i>call</i>	1
<i>set</i>	-	<i>carry</i>	1
Total	65	Total	86

Table 2: Distribution of collocations in unshifted [V NP PP] structure

	Swe	Nat	Total
Non-collocations	40 (62.5 %)	62 (72 %)	102
Collocations	24 (37.5 %)	24 (28 %)	48
Total	64	86	150

Overall, transitive [V NP PP] structures had a higher frequency rate in the native speakers'

speech than in the learners' (86:65). In raw frequency, *make*, *put* and *take* were the most frequent verbs for transitive structures in both the learner and the native data (a ratio of 15:9, 6:17 and 25:27 occurrences respectively).

One single instance of heavy noun phrase shift was found in the SW-LINDSEI data and none were found in the LOCNEC corpus. The single HNPS case is given below (the [V PP NP] sequence in bold font):

(1)

 or **wrote down on a piece of paper like . have**<?> **we got** <overlap /> **half**
<A> <overlap /> (uhu)
 a pint or . something like that

The PP *on a piece of paper* here precedes the shifted object NP *like have we got half a pint*. Note that the NP has the structure of an interrogative phrase (with subject-operator inversion, which is usually found in a question), even though it is incorporated in a larger declarative phrase. Furthermore, the NP is introduced with a discourse marker and a small pause. The verb collocates with a preposition (*wrote down*), and the preposition has therefore not been included in the following PP.

There are some cases of unshifted sentences in the Swedish learner data where the NP is longer than the following PP, which does not adhere to the end-weight principle. Some examples (only the NP in bold font):

(2)

 (er) and they had to take **an extra insurance** . for me<?> <laughs>

(3)

 <overlap /> well at any given moment in time you can find **at least three of those** in <overlap /> Jerusalem

Note that in these examples, the following PP is preceded by a hesitation, repetition of the initial word or small pause (.).

The native data also showed variation in the relative length of both NPs and PPs in unshifted order. Longer NPs were found in several cases:

(4)

I've got into the habit of bringing **a extra pair of trousers** with me if it's very wet and get changed when I get here [if I remember <\B>

(5)

 er .. I went with a school trip when I was teaching so we brought **a load of er [school kids** with us but they thought it was er <\B>

(6)

 all stuff like that we tried to put **as many things that they didn't understand** in the book so you know once they read it and kept reading it they'd kind of get the gist of it <\B>

In (6), the NP is also complex, since it contains the relative clause *that they didn't understand*. Note, however, that the NP is incomplete; the second *as* is omitted (the implied clause being *as many things as possible that they didn't understand*), which results in ellipsis.

There was one interesting case of a very long NP in the native data (the NP in bold font, followed by the PP):

(7)

you know they tried to put **the values of the southern sort of fairly cultured society** <\B>
<A> [mhm <A>
 [**in the south of England** on [to a northern area

In this utterance, the relative length of the two post-verbal constituents have a ratio of 15:6 words; the NP is almost three times the length of the PP. However, shifting did not occur. The speaker also showed hesitation when introducing the PP after such a long preceding phrase; there is a repetition (*a a*).

Instances where the PP is the longest phrase (in the relative sense) also adhere to the end-weight principle, since the phrase is placed in sentence-final position. Example from the learner data (the PP in bold font):

(8)

 yeah . it's a tragic love story <breathes> and I'm writing my C essay **about tragic love stories** now <overlap /> so I think it's

There were also instances of heavy prepositional phrases in the native data, for example:

(9)

[erm .. I got left some money **by a great uncle who died** <\B>

(10)

 so . at that time . I think it taught me a . an important lesson because at that time I'd made lots of plans **about what I was going [to do in the future** you know <\B>

In both these examples, the prepositional phrases contain embedded clauses, making them complex as well as long: *who died* in (9) and *what I was going to do in the future* in (10).

5.2 Dative Alternation

Below is the occurrence of the dative constructions found in the data (Table 2). The learner data is presented to the left and the native data to the right. The findings have been categorised as having prepositional order [V NP PP] (PO) or double object order [V NP_i NP_j] (DO) respectively for both corpora, and raw frequency rates as well as percentages are given for each of the three verbs. Further down, the number of cases where the direct object (Od) or indirect object (Oi) consist of a single pronoun is given for both constructions.

Table 3: Occurrences of dative alternation in ditransitive structures in both corpora (raw frequencies and percentage)

<i>Verb</i>	<i>Swe</i>			<i>Nat</i>		
	PO	DO	Total	PO	DO	Total
Buy	-	-	0	-	2 (100 %)	2
Give	7 (29 %)	17 (71 %)	24	5 (13 %)	35 (88 %)	40
Sell	2 (100 %)	-	2	-	1 (100 %)	1
Total	9	17	26	5	38	43

Table 4: Occurrences of single pronouns as direct and indirect objects (raw frequencies and percentage)

	<i>Swe</i>			<i>Nat</i>		
	PO	DO	Total	PO	DO	Total
Direct obj.	5 (55 %)	-	9	5 (100 %)	2 (5 %)	7
Indirect obj.	-	13 (76 %)	13	-	32 (84 %)	32

When the order was prepositional (PO), all five instances found in the native data had a single pronoun as direct object (Od). Five instances were also found in the learner data, but the corresponding percentage was only 55 %. Below are two examples of this type (direct object in bold font):

(11)

 <overlap /> people breeding them and selling **them** to anyone and they <breathes> so I had to . to learn about . these dogs as well so . they could be . sometimes very aggressive . not the dogs that I worked with <overlap /> they were

(12)

 giving **it** to: her future fiancé

Moreover, some equally long or longer PPs were found in sentence-final position as indirect objects also when the direct object was not a single pronoun:

(13)

 <overlap /> well it's fun because especially if you bring a friend .. and you walk around and you **try to . give a name to the<?> different things**

(14)

 (em) and if you watch T V shows . you you see a lot of intellectual programmes . and a lot of people watch them . and a lot of people read the magazines the newspapers and you discuss different topics .. (erm) . and the cultural life in itself (er) is (em) . <tuts> now I don't know the word .. (em) the state **gives a lot of money** <overlap /> **to the**
<A> <overlap /> yes yeah
 cultural life

In (13), the direct object consists of NP *a name*, which is not a single pronoun, and the following PP *to the different things* is twice as long (2:4). In (14), on the other hand, the constituents are equal in length. In neither case is any of the constituents complex.

For the double object order (DO), single pronouns dominated as indirect objects in both the native and the learner data (76 % in the learner data and 84 % in the native). Below there are two examples from the learner data:

(15)

 <overlap /> the way of thinking of like . the priority . they have on . what things are important it's not at all like here okay some things okay money's . im= important to have and gives **you** higher status but like their religion . a r= religious way of thinking I think

(16)

 and then they gave **me** a free bungalow I didn't have to pay any rent <overlap /> any

In the native data, the indirect object consisted of a single pronoun in all cases of DO except two (the Oi in bold font):

(17)

 in December <XX> give **children** presents

(18)

 [<laughs> <X> no said right er well we'll give **that one** a miss then

In both cases, the Od and Oi were simple and equal in length (1:1 for (17) and 2:2 for (18)). When the Oi was not a single pronoun in the DO construction, there was equal length between the two object constituents in the learner data (both objects in bold font):

(19)

 ma= (mm) give **people strength** <overlap /> people who like . can't really find their

(20)

 <breathes> and Lars Norén is so . good with with all these things because he . sort of (em) all of a sudden he gives **the audience a joke** <overlap /> <breathes>

In (19) the ratio is 1:1, and in (20), the ratio is 2:2.

There was one interesting case of reference in the native data where the Oi in a DO construction provided a new description for a previously mentioned entity. The whole paragraph is given for context (the entire DO construction in bold font):

(21)

 so she sits down again and then he paints what he doesn't see and he paints this . pretty woman who's smiling and looking friendly and . happy with life and she gets up and she's very very happy with that and she **gives the little man his money** and she goes back to show her friends and her friends are a little bit surprised because of course when stood next to the portrait they don't look quite the same

In this data, the native speaker made up a story around some pictures. Figuring in the story were a woman and a man, who painted two portraits of the woman. Thus, *he* in the data referred to the man, which was established at least three times prior to the marked part. The speaker then used *the little man* to refer to the same entity that had previously been described as *he*; providing a new description of an already given entity. This utterance differs from the other DO data in that shift is impossible (see chapter 6: Analysis and discussion).

6. Analysis and discussion

As seen above, only one instance of HNPS was discovered in the learner data and none at all in the native data. Clearly, the speakers preferred unshifted word order even in sentences with at least relatively heavy mid-positioned NPs. These results do not adhere to the end-weight principle, but rather appear to be evidence in support of earlier postulations (Stallings et al. 1998) that neither length nor complexity are necessarily always enough to trigger HNPS, not even when they occur together, as other factors may be involved.

There were some similarities between the learners' and the natives' speech: namely, preference for using double object order over prepositional order. The preference was stronger in the native data, a ratio of 17:9 in the learner data, and 38:5 for the native, although influence from the difference in size between the corpora cannot be ruled out. As Hawkins (1994) claims that the prepositional order is the underlying basic structure, it is interesting that there seems to be much fewer instances of this order. On the other hand, a change of structure into the double object order is very likely to occur if the (underlying) prepositional phrase is very short. Since most of the DO instances found were very short prepositional phrases, 76 % for learners and 86 % for natives (see examples 15 and 16, p. 21), these findings are well in line with such postulations. Again, the sizes of the corpora may have had some impact on the result. However, placing in final position a direct object that is longer than the indirect object does follow the end-weight principle.

Similarly, in the few instances of prepositional order that were found, the indirect objects were never single pronouns. This is expected as a longer indirect object would violate the end-weight principle if the prepositional *Od* were even slightly longer, which is likely to be the case since single pronouns make up the lightest possible constituent.

The fact that most of the PPs following heavy-NP sentences began with a hesitation or a small pause suggests that such constructions might be problematic for the speaker. A most likely possibility for this is that an internal heavy constituent is somehow disruptive to speech production. Even so, there appears to be a resistance toward shifting the challenge-inducing constituent to a more manageable final position. On the other hand, the NP needs to be significantly longer than the PP for a shift to occur Hawkins (1994). One can argue that in spoken language, any heavy element can pose a problem regardless of its position in the sentence because we are used to speaking in a more fragmented manner than in written

language. In order to draw conclusions on whether the speaker finds such sentence structures problematic, one would have to turn to other methods, for example analysis of the decision times of these sentences.

Interestingly, the only instance of HNPS (example 1, p. 18) also contained a pause before the shifted NP, suggesting that postponing the element to create a shifted order is not entirely beneficial from the speaker's point of view. While the speaker-oriented approach to weighted structures dictates that the speaker uses shifting to facilitate her own production through extended planning time of the forthcoming NP (Wasow 1997a), in this case, the outcome seems to be unsuccessful. Because the case in question is a single occurrence found in the learner data, however, the analysis is in need of corresponding HNPS in native data if any conclusions are to be drawn. One can argue that a listener-oriented approach is satisfied with this arrangement, since the listener gives a confirmation (*uhu*) at the beginning of the shifted NP, but as this is not a study of listener response in discourse, such a speculation will have to remain hypothetical.

The argument that disfluencies in speech are more prominent with reference to new discourse entities than given entities (Arnold et al. 2004) seems to confirm that new or previously unmentioned entities are less accessible. If the speaker pauses or hesitates in search for an appropriate word or description, that disfluency might, however, function as a signal for the listener that a reference to something previously unmentioned is going to take place; a cue to an introduction of a new discourse entity. It also seems reasonable to claim that new referents are more likely to be described in detail than given referents, and that more detailed descriptions are prone to be realised by syntactically heavier elements (*the little man* compared to *he* in 21). If so, then placing new information last is in concord with the tendency to place heavy elements last. Even Hawkins admits that "there are correlations between weight and information status" (Hawkins 1994: 215).

As previously mentioned in the results section, one native speaker's use of new description for an already established entity (example 21, p. 22) is very interesting. Even though the speaker feels the need to add information about the entity the last time it is mentioned, there can be no reasonable doubt regarding which entity is referred to; there is only one male figure in the discourse. Nowhere earlier in the interview is the size of the man commented upon; this is a new characteristic. Of course, it is possible that this particular case can be ascribed simply to irregularities in the speaker's speech, or to a desire to add to the story some colour that was spontaneously thought of at the last moment. If the prepositional order had been used for the same constituents instead, as in *gives his money to the little man*, *the little man* would most

likely have been interpreted as a new entity in the discourse, supported by the “new last” tendency in information distribution (Krifka 2008). In this case, the unshifted word order would result in a change of meaning which is inconsistent with the rest of the discourse, and so the double object order must be used instead. In other words, in example (21), the information principle imposes constriction on dative alternation.

Another factor that can be taken into account when considering the apparent lack of shifting in the material under study is the level of formality in the Swedish and native data. Since it is well established that the syntax of spoken language differs from that of written language on a number of points, it seems reasonable to assume that deviations from normal word order may differ as well. When Wasow (1997a; 1997b) studied the frequency rates of end-weight, he did so using written corpora. He stated that “[i]t seems plausible, however, that patterns of usage primarily motivated in spoken language would carry over to writing” (Wasow 1997a: 99), but this appears not to be the case for the speakers in this study, neither native nor learner, as their speech seems to be very resistant toward shifts dictated by the end-weight principle. Admittedly, in order to be able to determine similarities or differences between spoken and written languages of approximately the same group of speakers, one must conduct a study that compares spoken with written corpora.

As the Swedish learners used in this study are university students, they can be assumed to have a relatively high level of proficiency in English. Swedish and English are related languages on many areas, with some shared vocabulary and grammatical features, and as they are both head-final languages, it is only to be expected that they are both right-branching and display ‘light first, heavy last’ characteristics (Hawkins 1990). As earlier mentioned (section 2.3.2), acquisition is likely to be required more easily when the native language is similar to the target language (White 1990). Shared syntactic structures in Swedish and English may, at least partly, account for the similarities

While lexical bias and semantic influence have not always been ascribed much importance for the constructions treated in this study (Hawkins 1994), others have claimed that such “weight-sensitive phenomena cannot be attributed entirely to structural factors” (Wasow 1997a: 102). The results show a considerable difference in distribution of verbs in the cases of *find* and *put*, which were more frequent in the native data than in the learner data. Verbs may have a wider semantic range in the target language than in the native language. For example, the native speakers used *find* in constructions such as *I find it very difficult* (LOC 3071), of which there were fewer in the learners. In addition, the verb *put* has at least three counterparts in Swedish (*sätta, ställa, lägga*). Another possibility is that verbs may be used in contexts that

have no correspondence in the native language. If the target language uses a structure that is missing in the native language, one possible outcome is that the learner avoids that structure also in the target language (White 1990). *Put* is used only six times in the learner data, but 17 times in the native data. If the learners translated all three Swedish counterpart verbs, we would expect to find more instances of *put* in the learner data than in the native data, but as the case is the opposite, it could be an indication of avoidance.

7. Conclusions and further research

The literature on the subject of word order alternations reveal that there are different opinions on whether weight-sensitive phenomena ought to be ascribed mainly to syntactic conditions or to semantic and pragmatic principles. The results of this study, albeit small, did not show any consistent adherence to purely syntactic constraints that have been claimed to govern the occurrence of end-weighted structures. Length and complexity failed to predict shifted post-verbal clause constituent order, and heavy noun phrase shifting (HNPS) proved to be resistant to circumstances of such nature. However, a pragmatic motivation for HNPS could not be asserted either, since the data was so scarce, which leaves the possibility of a preceding consideration of the facilitation of speech production (or possibly listener comprehension) as the most probable. Naturally, further research must be made if this hypothesis is to gain more credibility.

Regardless, it is clear that deviations from basic word order are not an easy matter to handle, and that many other factors than those discussed here ought to be taken into account if a fuller picture of how Swedish learners handle weight-sensitive constructions is to be painted. For time-related reasons, this study has deliberately overlooked other end-weight phenomena such as subject extraposition and object complement extraposition, which could shed some light on the issue and point to overall trends or patterns in learners' construction of weight-sensitive sentences.

It is also crucial to consider how much importance should be attributed to information structure, reference comprehension and focus constraints as influences on clause constituent orders. Further research on the subject could preferably compare spoken and written language, on different levels of formality within the two registers as syntactic structure is a most flexible tool for communication.

References

- Arnold, Jennifer E. and Lao, Shin-Yi C. 2008. "Put in last position something previously unmentioned: Word order effects on referential expectancy and reference comprehension". *Language and Cognitive Processes*, 23: 2 (282-295).
- Bock, Kathryn and David Irwin. 1980. "Syntactic Effects of Information Availability in Sentence Production". *Journal of Verbal Learning and Behavior*, 19 (467-484).
- Bresnan, Joan, et al. 2007. "Predicting the dative alternation". *Cognitive Foundations of Interpretation* (69-94).
- Bresnan, Joan and Marilyn Ford. 2010. "Predicting syntax: Processing dative constructions in American and Australian varieties of English". *Language*, 86: 1 (168-213).
- De Cock, Sylvie. 2004. "Preferred sequences of words in NS and NNS speech". *Belgian Journal of English language and literatures (BELL)*, 2 (225-256).
- Gries and Newman. 2013. "Creating and using corpora". In: Podesva, Robert and Devyani Sharma (eds). *Research Methods in Linguistics*. Cambridge: Cambridge University Press.
- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 1990. "A Parsing Theory of Word Order Universals". *Linguistic Enquiry*, 21 (223-261).
- Hilgsmann, Philippe (ed.). 2014. "Fluency and disfluency in spoken English. Investigation of fluency profiles in English learner speech in comparison with native speech." *UCL: Université Catholique de Louvain*. Online at <http://www.uclouvain.be/en-470396.html>. Accessed 25th May 2015.

Krifka, Manfred. 2008. "Basic notions of information structure". *Acta Linguistica Hungarica*, 55: 3 (243-276).

Quirk, Randolph et al. 1972. *A Grammar of Contemporary English*. London: Longman.

Rasinger, Sebastian. 2010. "Quantitative Methods: Concepts, Frameworks and Issues". In: Litosseliti, Lia (eds.) *Research Methods in Linguistics*. London: Continuum International Publishing Group.

Rochemont, Michael and Peter Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.

Shih, Stephanie & Grafmiller, Jason (2011) 'Weighing in on end-weight'. Online at: <http://stanford.edu/~stephsus/ShihGrafmillerLSA2011.pdf>. Accessed March 30th 2015.

Siewierska, Anna. 1993. "Syntactic weight vs. information structure and word order variation in Polish". *Journal of Linguistics*, 29 (233-265).

Stallings, Lynne, Maryellen MacDonald and Pádraig O'Seaghdha. 1998. "Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift". *Journal of Memory and Language*, 39 (392-417).

Teubert, Wolfgang and Anna Čermáková. 2007. *Corpus Linguistics. A Short Introduction*. London: Continuum International Publishing Group.

Wasow, Thomas. 1997a. "Remarks on grammatical weight". *Language Variation and Change*, 9 (81-105).

Wasow, Thomas. 1997b. "End-weight from the speaker's perspective". *Journal of Psycholinguistic Research*, 26: 3 (347-361).

Wasow, Thomas and Jennifer Arnold. 2003. "Post-verbal constituent ordering in English". In: Rohdenburg, Mondorf (ed.). *Determinants of Grammatical Variation in English*, vol. 43. Walter de Gruyter.

White, Lydia. 1990. "Second language acquisition and universal grammar". *Studies in Second Language Acquisition*, 12 (121-133).