

Reflexivity in
English Prepositional Phrases
BA Thesis English Language and Culture
Utrecht University

Willeke S. van der Kallen

3220931

Supervised by Dr. Alexis Dimitriadis
and reviewed by Koen Sebregts

January 2013

Abstract

This paper is a corpus analysis of personal pronouns versus self-forms in American English locally bound prepositional phrases. Sentences containing a preposition immediately followed by a locally bound proform were taken from the tagged Brown Corpus (1979). For each locally bound prepositional phrase a number of semantic properties were indexed: meaningful, belief report, picture noun phrase, partitive, locative and/or directional. Statistical analyses showed that the choice of proform used per preposition varied significantly, and that the choice of proform used per semantic category as defined by Lederer (2009) (locative, directional and grammatical) varied significantly. These variations could not be attributed to any single semantic property. Future research should concentrate on the semantic categories per proform for each preposition.

Table of Contents

Abstract.....	2
Introduction.....	4
Theoretical Framework.....	7
Method.....	12
Results	16
Discussion.....	23
Conclusion	28
Works Cited.....	30
Sources	31
Appendix 1: brown-extract-PPL.py	32

Reflexivity in
English Prepositional Phrases

In recent years, there has been an increasing interest in grammatical reflexivity. When Noam Chomsky (1986) argued for the poverty of stimulus, stating that language is far too complex to learn considering the information children are subject to, his first example was based on the following two sentences:

- (1) I wonder who_i [the men_j expected to see them_j]
- (2) [the men_i expected to see them_j] (ibid, 8).

In sentence (1) the men would have expected that someone would come and see the men. Conversely, in sentence (2) the correct interpretation seems to be that the men expected to go and see someone else, e.g. the women. However, if the personal pronouns ‘them’ in these two sentences are replaced by reflexive pronouns, or ‘self-forms’ (König & Gast, 2002), their meanings change radically, as shown below.

- (3) I wonder who_i [the men_j expected to see themselves_i]
- (4) [the men_i expected to see themselves_i]

Now the self-form ‘themselves’ in sentence (4) indicates that the men had expected to look down at their hands, or maybe see a mirror, whereas sentence (3) leaves it entirely unsaid who actually saw the men or whom the men actually saw. Instead, sentence (3) speculates about some mysterious people who, according to the men’s expectations, were looking at themselves. As these four sentences show, there is a multitude of hidden rules governing the potential reflexivity and nature of reflexivity of personal pronouns and self-forms.

Many linguists have devoted themselves to exposing the rules of Universal Grammar (Chomsky, 1986) as they describe the potential of reflexivity, but much of the research up to now has been theoretical in nature (Ter Meulen, 2000; Reinhart & Reuland, 1993; König & Gast, 2002). Researchers have strived to identify and classify specific phenomena that

influence the syntactic and semantic properties of reflexivity (Kuno, 1987; Zribi-Hertz, 1989; Baker, 1995). As such, most papers regarding reflexivity consist of examples of syntactic structures and explanations of what differentiates these structures from other syntactic structures (König & Gast, 2002; Lees & Klima, 1963). While it is valuable to consider all the options and variations, this type of research serves to investigate what *can* more than what *does* happen. In apparent opposition to this trend, Haspelmath (2008) has attempted to explain phenomena in reflexivity not by theory alone, but by basing his explanations on frequencies of particular kinds of language use across various languages. Nevertheless, his work is mainly theoretical and uses the observed frequencies to explain why there are variations rather than investigate the variations themselves. Finally Lederer (2009) built a bridge between qualitative analysis, in which different phenomena receive equal attention regardless of their respective rarity, and quantitative analysis, which considers real-world frequencies to assess the relevance of various findings, in her “multi-method approach”.

The aim of this study is to closely examine language variation at work. As outlined in a recent study by Haspelmath (2008), in verb phrases there appears to be a strong inverse relation between how often a verb is used reflexively and how long the corresponding proform is. However, in prepositional phrases this pattern is not so prominent, and more variation in proform per preposition occurs. The current study will focus on prepositional phrases, in particular those in which the preposition is immediately followed by either a self-form or a personal pronoun. The main question investigated will be what determines whether a locally bound prepositional phrase is expressed via a self-form or a personal pronoun. A corpus study of the Brown Corpus allows for a number of theories to be subject to statistical testing. Haspelmath has suggested that, in the case of verb phrases, the more a verb is used with a locally bound proform, the shorter the proform is required to be to express reflexivity (2008, p. 47); this theory’s applicability to prepositional phrases will be tested. Furthermore,

Lederer's theory concerning spatial prepositional phrases (2009) shall be put to the test. The influence of optionality on reflexivity shall also be investigated. Finally, the choice of proform might be influenced by semantic structures rather than by specific prepositions. Specifically, this paper examines five semantic properties: spatial phrase (divided into locative and directional phrase), belief report, meaningfulness, picture noun-phrase, and partitive, each detailed below.

Theoretical Framework

When a personal pronoun refers to the same referent as a preceding pronoun or noun phrase within the same clause, it is usually replaced by a self-form (Mackenzie, 2007, p. 148). For example, some English self-forms are ‘myself’, ‘himself’ and ‘themselves’. Following König & Gasts definition of self-forms, this paper will use the term ‘self-form’ to encompass all words with an *x-self* or *x-selves* shape.

Recently, König and Gast (2002) have compiled one of the most complete classifications of self-forms to date. According to them, there are three categories of self-forms in English, only one of which can be counted as a reflexive. The other two word-classes would be logophor and intensifier. While reflexive self-forms refer to a noun phrase within the same clause, logophors are bound to a referent outside of the clause or are implied by the extra-lingual context of the sentence. Finally, intensifiers serve an emphatic purpose in grammatically unique positions. Intensifiers are the only one of the three classes of *self-forms* that do not take an argument position. Instead, they can appear directly after the corresponding noun phrase or at the end of the verb phrase and function as a synonym for ‘alone’, ‘specifically’ or ‘as well’.

In order to make these three word-classes applicable to personal pronouns as well as self-forms, the present study uses broader terms to express whether a proform is used reflexively or not. The criterion of reflexivity is whether or not a proform is locally bound. Therefore, the current paper will refer to locally bound proforms (comprised of reflexive self-forms and locally bound personal pronouns), not locally bound proforms (comprised of logophors and not locally bound personal pronouns), and proforms that are not bound at all (comprised of intensifiers and very rare occurrences of non-bound personal pronouns).

In König and Gasts classification, locally bound self-forms have been further divided into subcategories. The simplest of the set is the mandatory self-form. The mandatory self-

form can appear in an argument position, as part of a prepositional phrase that is in an argument position, or as the subject of a clausal phrase in that position. Its main function is to unify two entities within the same clause. For instance, consider the following three sentences.

(5) Fred_i fancies himself_i (König & Gast, 2002, p. 1).

(6) Fred_i fancies him_j.

In sentence (5) the subject and the object both refer to the person 'Fred'. If the self-form were to be replaced with a personal pronoun as in sentence (6), the meaning of the sentence would change to refer to two different people. That is why the self-form in this sentence is a mandatory self-form.

However, not all proforms are irreplaceable. Optional self-forms are distinct from their mandatory counterparts in that they can be replaced by personal pronouns while still being locally bound. Consider the following sentences.

(7) In the meantime, the King gathered an army around himself at Shrewsbury and then headed for Oxford [BNC ALK 203].

(8) The King gathered an army around him.

In the sentences (7) and (8) the personal pronoun 'him' and the self-form 'himself' both refer to 'the King', and are therefore locally bound. Because they can be swapped around without changing the referent, the proform is what König and Gast call an optional proform.

Although König and Gast (2002) describe more subclasses of locally bound self-forms as well as of unbound self-forms, the present study will not be examining these subclasses, and therefore they are not included in this chapter.

To explain the variation in proform choice for local binding in verb phrases per verb, Haspelmath (2008) propagated a sliding scale of 'introverted' and 'extroverted' verbs.

- (9) John dressed himself.
- (10) Nathan hit himself.
- (11) John dressed.
- (12) *Nathan hit.

In both (9) and (10), the self-form is used to refer to the subject of the clause. However, in sentence (11) the self-form can be omitted. In the words of König and Gast, there is alternation with zero. Implicitly, the sentence is still locally bound. Conversely, sentence (12) is ungrammatical. Haspelmath suggests that this is because ‘to dress’ is used more frequently in a locally bound context than in a not locally bound context, which makes ‘to dress’ an introverted verb. On the other hand, ‘to hit’ is rarely used in a sentence in which the object refers to the subject, so ‘to hit’ is an extroverted verb. Haspelmath theorises that introverted verbs can take the shortest form possible (such as ‘him’ or an omission of any type of proform) because people expect a locally bound referent, whereas extroverted verbs require it to be extra clear that in this case, the proform is locally bound. Such a frequency-based explanation might also apply to prepositional phrases. That is, there might be such a thing as ‘introverted prepositions’ and ‘extroverted prepositions’ influencing whether the locally bound proform needs to be a self-form or a personal pronoun.

Lederer (2009) indicates that the meaning of prepositional phrases influences proform choice in locally bound prepositional phrases. Of special interest to her was whether or not the prepositional phrase refers to location.

- (13) Mary told John about herself.
- (14) Mary pulled John toward herself.
- (15) Mary saw John near herself.

The prepositional phrase in (13) has nothing to do with location, and as such it is what Lederer calls a grammatical phrase. There are many kinds of grammatical phrases, and they

can perform a multitude of functions, as long as they have nothing to do with location.

Sentences (14) and (15) are locally bound spatial prepositional phrases. Their meaning relates to location respective to the co-referent. In the case of (14), John's location is changing. This is what Lederer calls a directional phrase. In the case of (15), however, the location of John is static. Lederer calls this a locative phrase. Certain prepositions, such as 'between', 'outside' and 'above' are ideally suited for locative use. Nevertheless, not all prepositions that can be used in such locative phrases are used exclusively in relation to location. For example, "before" can have a locative meaning ("in front of"), but also a temporal meaning ("at an earlier time than").

In addition to the previously discussed theories, a few interesting semantic properties have been isolated in the current research in order to compare frequencies of local binding with and without these properties. The first of these properties, meaningfulness, is alluded to in König and Gasts definitions of locally bound self-forms (2002). The current paper uses the term meaningfulness to describe whether or not the proform in question could be replaced by a different noun phrase while the sentence remains semantically sound. König and Gast suggest that this applies to both mandatory and optional self-forms, excluding the non-meaningful occurrences for being "inherently reflexive" (ibid).

Secondly, several attempts have been made to explain the behaviour of locally bound self-forms and personal pronouns as parts of noun phrases in which the noun itself is a depiction of the co-referent (Chomsky, 1986; Kuno, 1987; Lees & Klima, 1963). For instance, a photograph reflects whoever or whatever was in front of the camera at the time the picture was taken.

(16) John looks at a picture of himself.

Because a photograph reflects whoever or whatever was in front of the camera at the time the picture was taken, in sentence (16) John can see his own likeness on the photograph. Such depicting noun phrases are called picture noun phrases, or briefly picture-NP's.

Another semantic property that might play a role in locally bound prepositional phrases is “narrative point of view” (Zribi-Hertz, 1989) or ‘belief report’. In these phrases, the locally bound proform refers back to the person “whose speech, thoughts, or feelings are reported or reflected in a given linguistic context” (Clements, 1975, p. 141). In relation to such subjective viewpoints, Susumo Kuno considers various peculiarities in syntactic structures of sentences where “verbs such as *worry*, *disturb*, *please*, and *amuse* take a sentential subject” (1987, p.112). Such sentences require certain insider knowledge on the part of the speaker into the beliefs of the subject. In the present study, any sentences in which the prepositional phrase appears to stem from such insider knowledge was considered to be a belief report.

Finally, some noun phrases express a part of a whole using the preposition ‘of’ (Hoeksema, 1996), resulting in noun phrases like “some of them”, “two of you” or “a strange part of herself”. Such phrases are referred to in this paper as partitives.

Method

This study is based on the Brown corpus. This corpus, whose text samples were collected and tagged at Brown University, consists of American English prose that was originally published in 1961 (Kauhanen, 2011). It has been fully tagged by the Brown research team (*ibid*); every word has been classified according to its word-class so that subsequent researchers can single out specific grammatical structures as they occur in running prose. This makes the Brown corpus, and other tagged corpora like it (e.g. British National Corpus, Corpus of Spoken Professional American English, Archer corpus), ideally suited to aid researchers in investigating specific syntactic phenomena. In the current study, the Brown corpus was accessed through the use of NLTK (Bird, Loper & Klein, 2009).

First, all sentences containing a prepositional phrase in which a self-form immediately follows the preposition were gathered into a comma-separated values file by use of a small script written¹ in Python, as included in Appendix 1. Secondly, an alternate version of the same code was used to gather all sentences containing a prepositional phrase in which a personal pronoun follows the preposition into a second comma-separated values file. Prepositions are denoted in the Brown corpus with any tag starting with “IN”², self-forms are denoted with any tag containing “PPL”³, and personal pronouns, or more specifically objective personal pronouns, are denoted with any tag containing “PPO”⁴ [Brown]. The data from the two resulting tables was analysed manually.

¹ Courtesy of Dr. Alexis Dimitriadis

² Tags accepted as prepositions: “IN”, “IN-NC”, “IN-TL”, “IN-HL”, “IN+IN”, “IN+PPO”, “IN-TL-HL”

³ Tags accepted as self-forms: “PPL-TL”, “FW-PPL+VBZ”, “PPL-HL”, “PPL”, “PPLS”, “PPL-NC”, “FW-PPL”

⁴ Tags accepted as personal pronouns: “VB+PPO”, “PPO”, “PPO-NC”, “PPO-HL”, “IN+PPO”, “PPO-TL”, “FW-PPO+IN”, “FW-PPO”

Each sentence was indexed according to a set of criteria. Before anything else, the sentences were coded as either local or nonlocal. A sentence was marked as local if the proform referred to a noun phrase or pronoun within the same clause. Below are a few examples to illustrate what was considered to be local or nonlocal.

- (17) It_i can project long-range goals for itself_i. [Brown cg22 60]
- (18) At first the kitchen help_i was tolerant, but ordered their_i own supply of white bread for themselves_i. [Brown cf04 42]
- (19) The kitchen help ordered bread for the kitchen help.
- (20) The dancer_i who never loosens her_i hold on a parasol_j, begins to feel that it_j is part of herself_i. [Brown cg10 12]
- (21) The dancer feels [that the parasol is part of the dancer]

Sentence (17) is a fairly straightforward case of local binding. The verb has two arguments, the second of which is the prepositional phrase, and the proform inside it co-references with the subject. In example (18) there is a complex sentence structure that obfuscates the subject of the verb, but the simplified version of the relevant clause, sentence (19), reveals the local co-referencing. However, sentence (20) was not marked as locally co-referencing, because the proform co-references with the subject of a higher clause, as can be seen in (21). The current study focuses only on locally bound prepositional phrases, because that is where there might be a choice between a personal pronoun and a self-form. Any sentence that did not co-reference locally was excluded from further marking up.

Then each locally co-referencing sentence was checked to see whether it was an optional proform. It was considered optional if the proform could be swapped out for its counterpart (e.g. self-form swapped for personal pronoun) without changing the referent. When the optionality was not immediately apparent, the relevant part of the sentence with a

version for each proform was compared in Google Ngram Viewer [Google]. If both collocations appeared to be in use, the disputed sentence was marked as optional.

In the case of the non-optional proforms, it had to be considered whether it was a locally bound mandatory proform or not. A non-optional self-form might be an intensifier. A prepositional phrase that was synonymous to ‘alone’, ‘specifically’ or ‘as well’ was therefore marked as not bound. For the sake of completeness, non-optional personal pronouns were subjected to this same test.

Finally, a number of semantic properties were indexed for each of the remaining locally bound prepositional phrases. First, the phrases were checked on the ‘meaningful’ criterion. If the proform could be replaced by a noun phrase that did not express co-reference, and the sentence would still be semantically acceptable, it counted as meaningful. This was a largely subjective process. Then, the label of ‘belief report’ was attached to any locally bound prepositional phrase that revealed speech, thoughts, or feelings which required a specific narrative point of view. Next, all prepositional phrases that were part of a picture-NP were indexed, such as sentence (16) in the previous chapter. Additionally, any prepositional phrase that was part of a partitive construction was listed as such. Finally, following Lederer’s (2009) definition of spatial phrases, any locally bound prepositional phrases that conveyed location were marked as either locative or directional. If the location was static, the phrase was marked as locative, and if the phrase described movement relative to the location, it was marked as directional.

Once all the data were collected, they were subjected to statistical analysis. The dataset of locally bound self-forms was compared and contrasted with the dataset of their pronominal counterparts. To do so, several contingency tables (Peck & Devore, 2005, p. 586) were constructed. Following Peck and Devore’s explanation, each contingency table had a row for each possible value of the first category to be compared (e.g. each possible

preposition) and a column for each possible value of the second category to be compared (e.g. the proforms 'personal pronoun' and 'self-form'). For each combination of the two categories, the corresponding numbers of occurrences, or "frequencies" (ibid), were recorded. Finally, the marginal totals per row and column and the grand total were included in the contingency table.

Subsequently, each contingency table was subjected to either a Chi-square test for homogeneity (Peck & Devore, 2005, p. 589) or a Chi-square test for independence (ibid, p. 593). While mathematically identical, the Chi-square test for homogeneity compares sets of frequencies from two separately tested populations (e.g. self-forms and personal pronouns), whereas the Chi-square test for independence assesses the likelihood of an association between two variables in a single dataset (e.g. locative and directional phrases).

As detailed by Peck and Devore, the Chi-square test compares expected frequencies with observed frequencies, wherein the expected frequencies are estimated as a portion of the grand total frequency relative to the row total and column total (2005, p. 589). The Chi-square test for homogeneity results in a probability p that displays the likelihood that, if the two sets of frequencies' underlying category proportions are equal, any difference is due to chance. If this probability is suitably low ($p < 0.05$), it can be assumed with 95% certainty that the underlying category proportions are not equal.

Conversely, the Chi-square test for independence (Peck & Devore, 2005, p. 593) results in a probability p that displays the likelihood that, if the two variables are independent, any difference is due to chance. If this probability is suitably low ($p < 0.05$), it can be assumed with 95% certainty that the two variables are not independent from one another.

Results

Results were computed by establishing Boolean values for each of the following categories: local, meaningful, intensifier, optional, belief report, picture-NP, partitive, locative and directional. In total, 2327 sentences were analysed this way, including all uses of self-forms in prepositional phrases that could be found in the Brown Corpus (1979). The Brown Corpus contains 354 self-form prepositional phrases, of which 42 are optional self-forms and 158 are mandatory self-forms. The Brown Corpus also contains 4013 sentences with a preposition immediately followed by a personal pronoun. As many as 1973 of those sentences were included, chosen at random. Of the locally bound phrases in the pronominal list, approximately half were optional personal pronouns (100) and approximately half (93) were mandatory personal pronouns.

Table 1

Percentage of local use per proform

Proform	Local	Nonlocal	Total	% Local	% Nonlocal
Personal pronoun	193	1777	1970	10%	90%
Self-form	200	50	250	80%	20%
Total	393	1827	2220	18%	82%

Above, Table 1 shows the numbers of locally bound and not locally bound prepositional phrases per proform, as well as portraying their proportions in percentages. Not bound proforms have not been included in this table. As can be seen above, 90% of the personal pronoun phrases were not locally bound and had to be excluded from the bulk of the investigation. That is why far more personal pronoun phrases than self-form phrases had to be indexed in order to collect a comparable number of locally bound prepositional phrases. The large and statistically highly significant ($\chi^2 = 750.48$ with $df = 1$; $p < 0.001$) difference between the locality per proform is most likely due to their different grammatical properties. To illustrate, the following quote comes from an English grammar textbook: “Where a pronoun is coreferential with a noun phrase or pronoun mentioned earlier in the same clause,

it has to take a special form, that of a [self-form]” (Mackenzie, 2007, p. 148), which implies that all non-local co-reference needs to be expressed with a personal pronoun.

Not all prepositions had the same proportions of usage with personal pronouns and self-forms. Some were equally divided across both proforms, while some were skewed one way or the other. Statistical analysis showed that the spread of local personal pronouns across the various prepositions differed significantly ($\chi^2 = 80.05$ with $df = 9$; $p < 0.001$) from that of local self-forms. Shown below in Table 2 are the frequencies of personal pronouns and self-forms per preposition in locally bound prepositional phrases, sorted by the weighted percentage of self-form usage, and accompanied by the percentages of various syntactic and semantic properties. All percentages are relative to the amount of local uses of prepositional phrases with the relevant preposition, except for the local percentages, which are relative to the total amount of all bound uses of prepositional phrases with the relevant preposition.

Table 2
Properties per preposition in locally bound prepositional phrases

Preposition	Personal		Total	% Personal						Picture-		Meaningful
	pronoun	form		pronoun	form	Local	Locative	Directional	Optional	NP	report	
but	0	3	3	0%	100%	75%	0%	0%	67%	0%	67%	100%
about	3	15	18	17%	83%	15%	17%	0%	28%	11%	33%	89%
into	1	4	5	21%	79%	21%	0%	60%	40%	0%	20%	100%
for	9	35	44	21%	79%	18%	0%	0%	36%	0%	43%	100%
at	2	6	8	26%	74%	7%	0%	25%	13%	0%	88%	100%
upon	1	3	4	26%	74%	14%	0%	0%	50%	0%	25%	25%
to	18	51	69	27%	73%	17%	3%	4%	25%	0%	33%	86%
within	2	5	7	29%	71%	70%	100%	0%	100%	0%	29%	29%
after	1	2	3	34%	66%	23%	0%	33%	0%	0%	0%	67%
from	3	5	8	38%	62%	11%	0%	63%	38%	0%	38%	88%
on	3	4	7	44%	56%	9%	43%	29%	43%	0%	14%	86%
against	2	2	4	51%	49%	14%	0%	0%	50%	0%	75%	75%
by	1	1	2	51%	49%	5%	0%	0%	50%	0%	0%	100%
of	40	40	80	51%	49%	17%	6%	1%	19%	6%	29%	66%
over	1	1	2	51%	49%	9%	0%	100%	50%	0%	0%	100%
between	3	2	5	61%	39%	29%	40%	0%	60%	0%	40%	80%
in	15	8	23	66%	34%	26%	83%	0%	35%	0%	39%	43%
among	8	3	11	73%	27%	41%	73%	0%	27%	0%	0%	18%
with	27	9	36	76%	24%	18%	22%	44%	19%	0%	33%	36%
before	13	1	14	93%	7%	47%	64%	7%	64%	0%	29%	43%
above	1	0	1	100%	0%	10%	100%	0%	100%	0%	100%	100%
around	7	0	7	100%	0%	24%	100%	0%	100%	0%	43%	71%
behind	17	0	17	100%	0%	40%	88%	6%	82%	0%	18%	59%
below	2	0	2	100%	0%	67%	100%	0%	100%	0%	100%	100%
beneath	2	0	2	100%	0%	50%	100%	0%	100%	0%	100%	100%
beside	4	0	4	100%	0%	29%	100%	0%	50%	0%	50%	50%
beyond	1	0	1	100%	0%	14%	100%	0%	100%	0%	0%	100%
near	1	0	1	100%	0%	33%	100%	0%	100%	0%	0%	100%
nearest	1	0	1	100%	0%	50%	100%	0%	100%	0%	0%	100%
toward	2	0	2	100%	0%	13%	0%	100%	100%	0%	0%	100%
under	2	0	2	100%	0%	40%	50%	50%	100%	0%	50%	100%
Total	193	200	393	50%	50%	18%	26%	10%	36%	2%	34%	71%

It is apparent from this table that the frequency with which a locally bound prepositional phrase appears with a self-form rather than a personal pronoun cannot be attributed to any single factor. Extensive statistical analysis could find no equal category proportions between proform choice and any single one of the measured properties. However, a number of trends (statistically insignificant) do emerge from the data. For example, the prepositional phrases that only appeared in conjunction with personal pronouns also show a strong tendency to refer to static locations rather than directional movement, and are very often indexed as optional personal pronouns. Conversely, picture noun phrases and belief reports appear with prepositions that are more often used with self-forms.

Table 3
Lederer's categories examined

Category	Personal pronoun	Self-form	Total	% Personal pronoun	% Self-form
Locative	86	15	101	86%	14%
Directional	30	10	40	76%	24%
Grammatical	77	175	252	31%	69%
Total	193	200	393	50%	50%

Table 3 presents three categories from Lederer's (2009) investigation of locally bound prepositional phrases and their breakdown among self-forms and personal pronouns. The percentages are weighted to account for the different totals of personal pronouns and self-forms. There was a significant difference ($\chi^2 = 97.93$ with $df = 2$; $p < 0.001$) between the grammatical category, which appeared primarily with self-forms, and the two spatial categories, which used mostly personal pronouns.

Table 4
Lederer's non-grammatical categories examined

Category	Personal pronoun	Self-form	Total	% Personal pronoun	% Self-form
Locative	86	15	101	55%	45%
Directional	30	10	40	39%	61%
Total	116	25	181	50%	50%

The table above compares the locative and directional category as they appear in conjunction with self-forms and personal-pronouns, and once again adds weighted

percentages of self-form use versus pronominal use. Statistical analysis further provides support for Lederer's theory that the spatial categories differ significantly from one another in their co-occurrence with specific proforms ($\chi^2 = 11.44$ with $df = 1$; $p < 0.001$).

Another possible explanation for the variation in proforms per preposition lies in Haspelmath's (2008) frequentist approach. In relation to verb phrases, Haspelmath states that "[i]n all languages, verbs with higher frequency of [locally bound] use show shorter reflexive-marking forms than verbs with lower frequency of [locally bound] use" (ibid, p. 47). To test whether this theory is applicable to prepositional phrases, Table 2 (above) was used to compare the percentages of local use per preposition with the percentages of locally bound personal pronouns versus self-forms. Non-bound proforms have been excluded from the calculations. The data appears to reveal a slight tendency for prepositions that are often used with a locally bound proform to express that locality with a personal pronoun rather than a self-form, but this tendency was not statistically significant.

Figure 1

The percentage of local prepositional phrases compared with the percentage of personal pronouns among local prepositional phrases

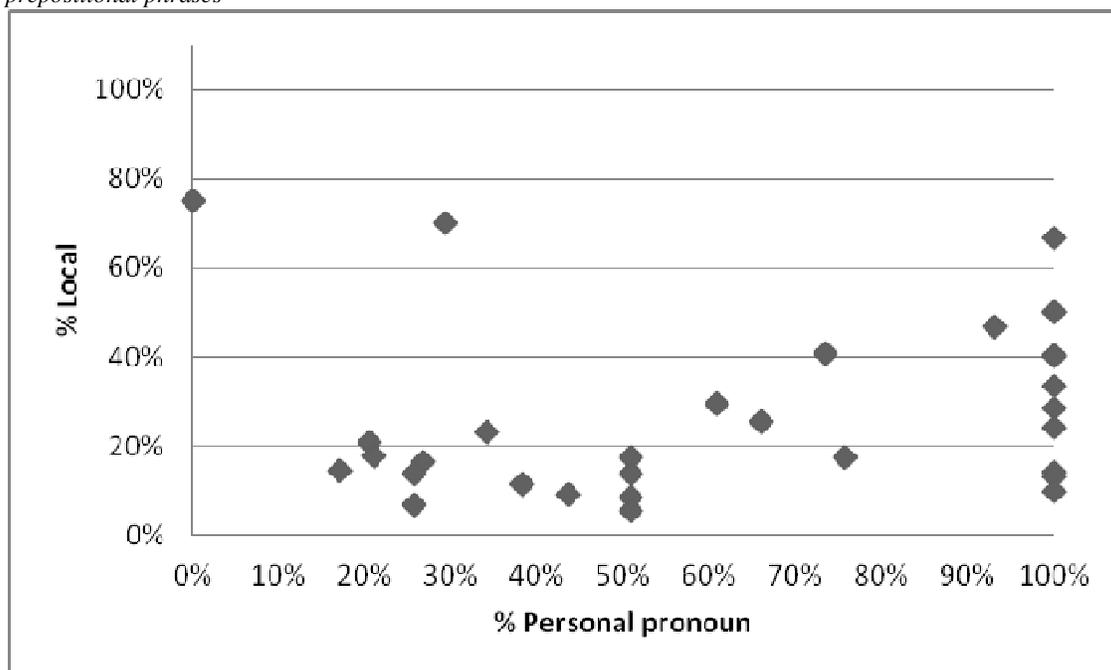


Figure 1 above compares the frequencies of local use per preposition with the percentage of local pronominal use as opposed to local self-form use, and makes the extent of their relationship visible. The two outliers are ‘but’ with 75% (3/4) locality and no pronominal uses, and ‘within’ with 70% (7/10) locality and 29% (2/7) pronominal uses. Haspelmath’s (2008) frequentist explanation of reflexivity in verb phrases, which poses that the percentage of local use of the verb will have an inverse relationship with the length of the proforms used to express local binding, cannot be statistically proven to be applicable to individual prepositional phrases. However, it may still be part of why prepositions have varying proportions of co-occurrence with locally bound self-forms and personal pronouns.

One final property that may play a key role in explaining the variation between personal pronouns and self-forms in locally bound PP’s is the difference between mandatory proforms and optional proforms. As noted above in relation to Table 2, there appears to be a correlation between optionality and proform choice. An initial Chi-square test showed significant differences between the breakdowns of mandatory and optional proforms per preposition ($\chi^2 = 60.01$ with $df = 8$; $p < 0.001$), which means that there is an association between specific prepositions and optionality.

Table 5
Optionality of locally bound self-form phrases per preposition

Preposition	Optional	Mandatory	Total	% Optional
before	1	0	1	100%
within	5	0	5	100%
but	2	1	3	67%
upon	2	1	3	67%
between	1	1	2	50%
among	1	2	3	33%
for	10	25	35	29%
in	2	6	8	25%
into	1	3	4	25%
on	1	3	4	25%
from	1	4	5	20%
of	6	34	40	15%
to	7	44	51	14%
about	2	13	15	13%
after	0	2	2	0%
against	0	2	2	0%
at	0	6	6	0%
by	0	1	1	0%
over	0	1	1	0%
with	0	9	9	0%
Total	42	158	200	21%

Shown above is a table listing the frequencies of optional and mandatory self-forms per preposition. The Chi-square test did not show any significant differences between the breakdown of optional self-forms among various prepositions and the spread of mandatory self-forms among those same prepositions ($\chi^2 = 4.68$ with $df = 3$; $p > 0.100$). Therefore, there is no association between preposition and optionality among self-forms, unlike there is for all proforms viewed together.

Table 6

Optionality of locally bound personal pronoun phrases per preposition

Preposition	Optional	Mandatory	Total	% Optional
about	3	0	3	100%
above	1	0	1	100%
against	2	0	2	100%
around	7	0	7	100%
below	2	0	2	100%
beneath	2	0	2	100%
beyond	1	0	1	100%
by	1	0	1	100%
into	1	0	1	100%
near	1	0	1	100%
nearest	1	0	1	100%
over	1	0	1	100%
toward	2	0	2	100%
under	2	0	2	100%
within	2	0	2	100%
behind	14	3	17	82%
between	2	1	3	67%
for	6	3	9	67%
from	2	1	3	67%
on	2	1	3	67%
before	8	5	13	62%
to	10	8	18	56%
at	1	1	2	50%
beside	2	2	4	50%
in	6	9	15	40%
with	7	20	27	26%
among	2	6	8	25%
of	9	31	40	23%
after	0	1	1	0%
upon	0	1	1	0%
Total	100	93	193	52%

Table 6 provides the frequencies of optional and mandatory personal pronouns per preposition. A Chi-square test for homogeneity showed that there are significant differences in the proportions of optional versus mandatory personal pronouns among the prepositions ($\chi^2 = 43.04$ with $df = 6$; $p < 0.001$). This suggests that in prepositional phrases containing locally bound personal pronouns, the specific preposition influences whether the locally bound personal pronoun is an optional or mandatory personal pronoun. Among the

prepositions with high enough⁵ frequencies to be measurable as individual categories, the main outliers appear to be the prepositions ‘of’ and ‘with’, which are used mostly with mandatory personal pronouns rather than optional personal pronouns. In the case of the preposition ‘of’, this discrepancy might be explained by the high number of partitive phrases. These phrases only appear with the preposition ‘of’, and account for 74% (23/31) of the ‘of’ mandatory proforms, as opposed to 33% (3/9) of the ‘of’ optional proforms. In the case of the preposition ‘with’, the high proportion of mandatory proforms might be attributed to the high proportion of directional uses, which account for 59% (16/27) of the ‘with’ phrases. However, it is unclear whether there is any association between directionality and optionality.

⁵ If the expected cell count⁶ in each cell was 5 or higher, the frequencies in that row were considered high enough.

⁶ (row total * column total) / grand total

Discussion

This study set out to examine what guides the choice between self-forms and personal pronouns in locally bound prepositional phrases in which the preposition is immediately followed by any of the two proforms. Statistical analysis of over two thousand sentences from the Brown Corpus (1979) revealed that approximately 10% of all bound prepositional personal pronoun phrases were locally bound, while as many as 80% of all bound prepositional self-form phrases were locally bound. A Chi-square test for homogeneity showed that the ratio of personal pronouns versus self-forms in locally bound prepositional phrases is not the same for each preposition. Furthermore, no single deciding factor could be identified among the properties investigated that could account for the variations in this ratio. There are, however, a number of factors that may play a part in proform choice.

In 2002, König and Gast suggested that optional proforms only occur where an argument position is filled by “the whole prepositional phrase, which has either a [locative] or directional meaning” (2002, p. 5). They posed this in reference to Reinhart and Reuland (1993), who investigated those two types of locally bound optional prepositional phrases, causing the faulty impression that these types are the only types of locally bound optional prepositional phrases. Consider, however, the following sentence.

- (22) If our sincerity is granted, and it is granted, the discrepancy can only be explained by the fact that we have come to believe hearsay and legend about ourselves in preference to an understanding gained by earnest self-examination.
[Brown cg29 48]

From sentence (22) and 58 other examples encountered in the data it can be concluded that there are various other types of optional reflexive prepositional phrases that are neither locative nor directional in meaning. These findings are in keeping with Lederer’s (2009) theory, which classifies non-spatial phrases as grammatical phrases.

As mentioned above, Lederer (2009) splits locally bound prepositional phrases in two categories: spatial and grammatical. Spatial phrases refer to location, whereas grammatical phrases do not. The spatial phrases have been further divided into locative and directional phrases, in which locative phrases refer to static locations whereas directional phrases refer to movement. Two consecutive Chi-square tests for homogeneity confirm her theory that grammatical phrases have different proportions of personal pronouns versus self-forms than the two categories of spatial phrases, and that the two categories of spatial phrases also vary significantly from one another. Grammatical phrases appear mostly in conjunction with self-forms, whereas spatial phrases appear primarily with personal pronouns. Among spatial phrases, locative phrases have a higher ratio of personal pronouns than directional phrases. These results corroborate Lederer's theory (2009).

Only 14% (25/181) of the locally bound spatial phrases contained a self-form. What is surprising is that 56% (14/25) of the spatial self-form phrases refer to a location inside the locally bound referent or between the locally bound referents, as in the following three sentences.

- (23) Visually, these approximated what he was feeling within himself. [Brown ck14 28]
- (24) I forced confidence into myself. [Brown ck29 47]
- (25) If the authorities remarked on her curious avocation, they did so among themselves. [Brown cm05 91]

Sentence (23) contains a locative phrase, and refers to a feeling situated inside the subject. In example (24), the directional phrase describes a movement towards the inside of the subject. Finally the locative phrase in sentence (25) refers to a location somewhere between the various members of 'the authorities'. The discrepancy in meaning between pronominal spatial phrases and self-form spatial phrases is also reflected in Table 2 above, which shows that

while most of the exclusively-spatial prepositions occur only with personal pronouns, the exclusively-spatial preposition ‘within’ only appears with a personal pronoun in 29% (2/7) of the cases. The other 44% (11/25) locally bound spatial self-form phrases might have a self-form because of emphasis placed on the referent.

In relation to the frequentist theory as adapted from Haspelmath (2008), this study has been unable to demonstrate that the frequency with which a preposition is used with a locally bound proform as opposed to a non-locally bound proform has any significant correlation with the choice of which proform is used. Although not significant, there did appear to be a slight tendency for the more ‘introverted’ prepositions to appear with a personal pronoun. It might be the case that how introverted a preposition is does not play a deciding role in the proform choice, but does influence this choice in a lesser capacity. A possible explanation for this might be that the locality of introverted prepositions does not need to be emphasised as much to be clearly recognisable as locally bound, removing the necessity for the use of a self-form.

One unanticipated finding was that the prepositions that appeared exclusively in locally bound pronominal PP’s still featured 88% (35/40) optional personal pronouns. These rather contradictory findings may be due to the amount of emphasis placed on the prepositional phrase.

(26) Phil shut the door behind him. [Brown cp24 190]

(27) I heard the screech of brakes behind me, an insane burst of laughter beneath me. [Brown cn16 116]

(28) Phil shut the door behind himself.

(29) I heard the screech of brakes behind John, an insane burst of laughter beneath myself.

Sentences (26) and (27) contain three of the 35 examples of optional personal pronoun phrases that appear with prepositions that never occurred with locally bound self-forms. In the case of (26), swapping the proform as in (28) would still permit co-reference to Phil, but would identify the door that Phil shut. Where there is only one door in sentence (26), there are at least two in (28) so it needs to be specified which door Phil shut. Sentence (27) is a more complex example, and its variation (29) emphasises the identity of the subject rather than the object.

Regarding optional proforms in general, a possible explanation for the regularity of optionality in prepositional self-form phrases lies in the predominantly local use of self-forms. Over 50% of prepositional self-form phrases (including the unbound intensifiers) are used locally, which may have made locally bound prepositional self-form phrases so common that over the years the proportions of optionality of these phrases evened out. As noted by Bouma and Spenader, the fact that self-forms are optional more often than personal pronouns might also be attributed to the amount of emphasis placed on the reflexive (2009, p. 122). The varying proportions of optionality per preposition in pronominal PP's on the other hand might be caused by semantic properties that were not indexed in the current research, such as demonstrated below.

(30) He did it for himself.

(31) She told him about herself.

Sentence (30) is a benefactive construction (Bosse, Bruening & Yamada, 2012, p. 1186), and example (31) explanation of content, neither of which have been examined closely in the present study. However, the relatively small sample size of locally bound proforms has caused the majority of the prepositions in both prepositional personal pronoun phrases and prepositional self-form phrases to be grouped together as 'other' in the statistical analyses, so these results should be interpreted with caution.

On a final note: while the present study has included meaningfulness as a variable, König and Gast have excluded all non-meaningful cases from their consideration (2002, p. 3). As such, they were able to claim that every locally bound self-form “occupies a syntactic position that could also be taken by some other noun phrase” (ibid). Such claims are unsatisfactory because they are true by definition, rather than a result of the available data. The present study includes 127 non-meaningful locally bound proforms.

The current research has found significant evidence in support of Lederer’s (2009) classifications of locative, directional and grammatical locally bound prepositional proform phrases. In addition, there was some evidence to suggest that internal locations play a role in proform choice among spatial phrases. Conversely, the results of the current study do not support a frequentist classification of introverted versus extroverted prepositions. These findings have important implications for developing a reliable model of reflexivity in English prepositional phrases, and may help to focus future studies on the guiding principles of proform choice in locally bound prepositional proform phrases.

Conclusion

The present study was designed to determine the relationship between specific prepositions in locally bound prepositional proform phrases and which proform is used. To do so, 354 self-form phrases and 1973 personal pronoun phrases from the Brown Corpus (1979) were indexed according to word-type and several semantic properties. These properties were: meaningful, belief report, picture-NP, partitive, locative and directional. The resulting data were then submitted to statistical analysis, which revealed that there are significant differences in the proportion of local personal pronoun phrases versus local self-form phrases among Lederer's (2009) locative, directional and grammatical categories. Furthermore, internal locations appeared to influence proform choice for both locative and directional phrases. However, no significant evidence was found to suggest that Haspelmath's (2008) distinction between introverted and extroverted verbs is applicable to prepositions.

The Brown Corpus contained a relatively low number of reflexive prepositional phrases, estimated at a total of 600 reflexive prepositional phrases. In investigating the data per individual preposition, many had too few occurrences to be examined individually, and instead these occurrences were grouped together under a category of "other prepositions". Therefore, the patterns found using the statistically relevant prepositions might not apply to the less commonly found prepositions. Furthermore, the Brown Corpus was compiled from texts published in 1961, which is over 50 years ago (Kauhanen, 2011). The frequencies observed in texts from 1961 are not necessarily applicable to present day English. Additionally, the Brown Corpus was compiled from edited American English prose (*ibid*). This means that these findings cannot be extrapolated to varieties of English other than American English, nor can they be extrapolated to spoken language. There is also little information available as to the influence of editing on reflexive prepositional phrases. Finally, the present study has only considered reflexive prepositional phrases in which the preposition

was immediately followed by a self-form or personal pronoun, excluding an unknown amount of more complex reflexive prepositional phrases.

This research has thrown up many questions in need of further investigation. Most importantly, this research has failed to determine a single factor that decides the proform choice per preposition, despite proving that the proform choice differs significantly among the various prepositions. What is now needed is a cross-category analysis of variance (ANOVA) to determine whether all prepositions vary significantly from one another in proform proportions, or whether certain prepositions present anomalies that require specific attention. Additionally, more criteria could be added to the indexation to test other theories for these irregularities. The reliability of the ‘meaningfulness’ criterion could be improved by a focus group of native speakers of English. More broadly, research is also needed to determine whether all varieties of English exhibit the same proportions. For instance, data from this Brown Corpus study could be compared with a study of the British National Corpus (BNC). It could also be interesting to compare the results of the current research with a similar investigation of the Freiburg-Brown corpus of American English (or Frown), which consists of texts from 1992 (Tyrkkö, 2012), in order to explore potential variations over time. Finally, further work needs to be done to assess the proportions of proform choice in grammatical contexts other than prepositional phrases in which the proform immediately follows the preposition.

Works Cited

- Baker, C. L. (1995). Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English. *Language* 71.1: 63-101.
Retrieved from <http://www.jstor.org>
- Bird, S., Loper, E. & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- Bosse, S., Bruening, B., & Yamada, M. (2012). Affected experiencers. *Natural language and linguistic theory* 30.4: 1185-1230. Retrieved from <http://link.springer.com>
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Westport, CT: Greenwood. Retrieved from <http://books.google.nl>
- Peck, R. & Devore, J. L. (2005). *Statistics: the exploration and analysis of data*. 7th international ed. Belmont, CA: Brooks/Cole.
- Haspelmath, M. (2008). A frequentist explanation of some universals of reflexive marking. *Linguistic Discovery* 6.1:40-63. Retrieved from <http://www.eva.mpg.de>
- Hoeksema, J. (1996). Introduction. In J. Hoeksema (ed.), *Partitives: Studies on the Syntax and Semantics of Partitive and Related Constructions* (pp. 1-24). Berlin: De Gruyter.
Retrieved from <http://scholar.google.nl>
- Kauhanen, H. (2011, March 20). *The Standard Corpus of Present-Day Edited American English (the Brown Corpus)*. Retrieved from the Research Unit for the Study of Variation, Contacts and Change in English website: <http://www.helsinki.fi/varieng/>
- König, E. & Gast V. (2002). Reflexive pronouns and other uses of self-forms in English. *Zeitschrift für Anglistik und Amerikanistik* 50.3: 1-14.
- Kuno, S. (1987). *Functional syntax: Anaphora, discourse and empathy*. Chicago: Chicago University Press.

Lees & Klima. (1963). Rules for English pronominalization. *Language* 39.1: 17-28.

Retrieved from <http://www.jstor.org>

Mackenzie, J. Lachlan. (2007). Principles & pitfalls of English grammar. Bussum: Coutinho.

Reinhart T. & Reuland E. (1993). Reflexivity. *Linguistic Inquiry* 24.4: 657-720. Retrieved

from <http://www.jstor.org/>

Ter Meulen, A. G. B. (2000). Optimal reflexivity in Dutch. *Journal of Semantics* 17: 263-280.

Retrieved from <http://joi.oxfordjournals.org/>

Tyrkkö, J. (2012). The Freiburg-Brown corpus of American English (Frown). Retrieved from the Research Unit for the Study of Variation, Contacts and Change in English website:

<http://www.helsinki.fi/varieng/>

Zribi-Hertz, A. (1989). Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 64.4: 695-727. Retrieved from

<http://www.jstor.org/>

Sources

[BNC] British National Corpus.

[Brown] A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown). 1964, 1971, 1979. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island.

[Google] Google Ngrams Viewer. Accessed at <http://books.google.com/ngrams>.

Appendix 1: brown-extract-PPL.py

```

"""
Find sentences matching our pattern and write them out to a file
"""

import nltk
from nltk.corpus import brown
from nltk.util import bigrams

def prepare_sentence(sent, matchid):
    """Convert the sentence into a printable string of words,
    capitalizing the pair of words that starts at index `id`

    >>> prepare_sentence([('an', 'AT'), ('eye', 'NN'), ('for', 'IN'), ('change', 'NN') ], 1)
    'an EYE FOR change'
    """

    # Extract the words, discarding the tags.
    words = [ w[0] for w in sent ]
    # Capitalize the two words at the indicated index
    words[matchid] = words[matchid].upper()
    words[matchid+1] = words[matchid+1].upper()

    return " ".join(words)

# `columns` must contain the same keys as `makerow()` creates.
columns = "No Filename Position Sentence Preposition Proform Type Locality".split()

def makerow(sent, count, fid, sent_id, word1, word2, kind):
    """Prepare a dictionary containing a row to write to the CSV file"""

    row = { 'No': count,
            'Filename': fid,
            'Position': sent_id,
            'Sentence': sent,
            'Preposition': word1,
            'Proform': word2,
            'Type': kind,      # pronoun or reflexive
            'Locality': "",   # to be coded manually
            }

    return row

```

```
# Collect all matching sentences, with metadata, into an array of dictionaries
# We search one file at a time, so we know which file we're in.
# We also save the index of the sentence in the file (`sent_id`), so it can
# be uniquely identified.
```

```
sentences = [ ]
count = 1 # This is just the row number, as in the manual list
for fid in brown.fileids():
    for sent_id, sent in enumerate(brown.tagged_sents(fileids=fid)):
        # progress marker: Print a dot every thousand sentences
        if sent_id % 1000 == 0:
            print ".",

        for word_id, ((p, tag1), (refl, tag2)) in enumerate(bigrams(sent)):
            if tag1.startswith('IN') and "PPL" in tag2:
                cleansent = prepare_sentence(sent, word_id)
                row = makerow(cleansent, # Sentence text
                              count, # index in the results list
                              fid, # filename
                              sent_id, # index of sentence in the file
                              p, refl, # the matched words
                              "Reflexive" # "Reflexive" or "Pronoun"
                              )
                sentences.append(row)
            count += 1
```

```
# Write the list of dictionaries to a file
import csv
csvfile = open("matches.csv", "w")
writer = csv.DictWriter(csvfile, columns, dialect='excel', lineterminator='\n')

writer.writerow(dict(zip(columns,columns))) # write out the headers as the first row
writer.writerows(sentences)
csvfile.close()
```