https://doi.org/10.31820/f.31.1.11

Jakob Lenardič and Gašper Ilc

ENGLISH RAISING PREDICATES AND (NON-)FINITE CLAUSES: DIACHRONIC AND SYNCHRONIC PERSPECTIVES

Jakob Lenardič, Faculty of Arts, jakob.lenardic@ff.uni-lj.si, Ljubljana
Gašper Ilc, PhD, Faculty of Arts, gasper.ilc@ff.uni-lj.si, Ljubljana

pregledni članak
UDK 811.111'367.3
rukopis primljen: 14. listopada 2018.; prihvaćen za tisak: 4. srpnja 2019.

In this paper, we present a diachronic and synchronic analysis of raising and extraposition constructions in the historical Brown Corpus and the more contemporary English Web Corpus 2015. *We begin by establishing two diachronic facts: first, raising constructions are used much more frequently than their semantically equivalent extraposition variants, and second, the distribution of raising and extraposition remains – rather exceptionally in comparison to other structures allowing for finite/non-finite variation – diachronically consistent from the beginning of the 20th century to 2015. We then supplement this unique diachronic distribution with an analysis of the most recent corpus data, which shows that the choice between the two semantically equivalent constructions is governed by distinct structural factors unique to each construction. Concretely, we show that the raising construction is frequently used as a relative clause, whereas the extraposition variant generally resists such a syntactic role. By contrast, we show that a prominent factor in favour of extraposition relates to the negative marker, which is placed with similar frequency both in the matrix and in the embedded clause of the extraposition construction in contrast to the raising variant, which uses the negative marker almost exclusively in the matrix clause. Lastly, we show that extraposition constructions contain modal verbs in the matrix clause more frequently than the raising variants and we tie this observation to the idea that the clausal composition of the extraposition construction is structurally more suited for expressing tentativeness.*

Key words: *non-finite clauses; extraposition; raising; corpus linguistics; formal linguistics; diachronic change*

1. Introduction

In this paper, we present a corpus-based analysis of the raising and extraposition constructions, exemplified by the sentence pair in (1) and (2).[1]

(1) *George is likely to have committed a crime.* (raising)

(2) *It is likely that George (has) committed a crime.* (extraposition)

The two constructions are interesting because they present a uniquely complex structural environment in which the use of finite and non-finite clauses alternates. The raising construction in (1), which contains an embedded non-finite clause, involves cross-clausal constituent movement (cf. Postal 1974 for one of the earliest analyses). The NP *George*, which is thematically an argument of the predicate *committed,* is said to *raise* from its initial position as an underlying constituent of the embedded clause to the subject position of the matrix clause, as shown in (3a). This is done for reasons that are entirely formal (in the sense that *George* does not become the semantic subject of the predicate *likely* post-raising), and failure to do so results in an ungrammatical sentence (3b), since a non-finite clause is in comparison to a finite one structurally poorer and therefore cannot host an NP as its grammatical subject.

(3) a. *George is likely* [*to have ~~George~~ committed a crime*.]

b. **It is likely George to have committed a crime*.

[1] Note that in generative grammar, sentences (1) and (2) are not understood as derivationally related constructions; that is, neither is a transformational variant of the other, nor are they derived from a separate underlying structure such as *That George has committed a crime is likely*. On the one hand, the syntactic difference (i.e., the raising vs. extraposition distinction) between sentences (1) and (2) follows from the fact that the embedded clause in (1) contains a tense projection that is impoverished, i.e., "*φ*-incomplete", which triggers the displacement of the NP *George* from its initial position in the embedded clause to the subject position of the matrix clause, whereas the embedded clause in sentence (2) does not trigger displacement because it is finite – i.e., it contains a tense projection that is *φ*-complete; see Section 3.3.1 for details. In other words, sentences (1) and (2) are the results of separate syntactic derivations. On the other hand, complementation in English – be it clausal or phrasal – follows the predicate that selects for it. In sentence (2), this means that the observed position of the embedded clause *that George (has) committed a crime*, which follows the selecting predicate *is likely*, is the original ("deep structure") position in the syntactic derivation, so the sentence *That George has committed a crime is likely* is actually the derived variant, involving displacement of the embedded clause to the sentence-initial position; see Radford (2004: 266–267) for further discussion.

By contrast, the embedded clause in sentence (2) is finite and is not structurally impoverished.[2] Consequently, the NP *George* is allowed to (and in fact must) remain as the embedded grammatical subject, whereas the matrix subject position is filled by the pronoun *it*, which is semantically vacuous and obligatorily enters the structure to satisfy the formal requirement which states that a non-*pro*-drop language like English always has to have a phonologically overt NP in the grammatical subject position in finite clauses. In this respect, the term *extraposition* refers to the fact that, unlike in the raising variant, all the constituents of the clausal complement follow the predicate that semantically selects for it (Radford 2004: 266–268).

Set against the generative theoretical background, the aim of the paper is to analyse how the alternation manifests itself in actual usage, both diachronically and synchronically. We first present distributional data from the historical *Brown Family* corpus and the contemporary *English Web Corpus 2015* with which we establish two diachronic facts: on the one hand, the raising construction is used much more frequently than the semantically equivalent extraposition construction from the beginning of the 20th century to 2015, while on the other, the distribution of the two constructions remains diachronically stable (i.e., unchanging) within the same time period; that is, the use of non-finite raising variants does *not* become more frequent at the expense of finite extraposition variants in the observed time period. Crucially, this latter lack of change in the distribution is inconsistent with the diachronic generalisation which states that "the spread of infinitival complement clauses at the expense of finite ones is a phenomenon which, in one form of another, can be observed throughout the whole recorded history of English" (Leech et al. 2009: 185).

[2] The term *impoverished* is used to refer to the fact that non-finite clauses cannot host grammatical subjects because they do not display *φ*-feature agreement and consequently cannot assign structural case to the embedded subject position (e.g., Chomsky 2001; for an alternative proposal, see Pesetsky and Torrego 2017, who tie the ban on NPs in subject position to the fact that non-finite clauses lack an interpretable tense feature). Such agreement can be otherwise observed in the fact that the NP *George* obligatorily triggers the third person inflection *-s* on the auxiliary verb in the finite variant (2). Note, however, that a sentence like *It is likely for George to have committed a crime* is fine because *George* receives case not under *φ*-feature agreement, which does not take place in infinitival clauses in any case, but from the preposition *for*; crucially, prepositions are taken to be case-assigning constituents in generative syntax (Radford 2004: 314). We discuss the impoverished structure of non-finite clauses in detail in Section 3.3.1.

The synchronic analysis, conducted on the basis of data from the contemporary *English Web Corpus 2015*, supplements our diachronic data by identifying several structural factors that are tied to the complex inter-clausal syntactic dependencies underlying the alternation. Such factors show that the two constructions are not equally suited for all structural contexts, and they show that there are cases in which the extraposition construction serves as a better way of encoding information than the raising construction. This, in turn, also provides an explanation as to why the raising construction has not become diachronically more frequently used at the expense of extraposition.

The paper is structured as follows. Section 2 presents the diachronic analysis of the two constructions. Section 3 presents the synchronic analysis. Section 4 is the conclusion.

2. The Diachronic Distribution of Raising and Extraposition

2.1. Extracting the data

In this section, we present the diachronic distribution of the raising and extraposition construction for three raising predicates – the verbs *appear* and *seem* and the adjective *likely*.[3] We have first surveyed the historical distribution of the raising and extraposition constructions on the basis of the *Brown Family*, which is a 6.7-million-word corpus of American and British English divided into 4 subcorpora at roughly 30-year intervals. These are the 1931, 1961, 1991 and 2006 subcorpora, each of which is roughly equivalent in size (between 1 and 2 million words). Since the corpus is accessible through

[3] We have chosen these predicates because they constitute a semantically uniform group, in the sense that they θ-mark (i.e., thematically select) only the proposition denoted by the embedded clause. In this respect, they differ from predicates like *expect*, which can also occur in raising and extraposition constructions when used in the passive voice, as in the following two sentences:

(i) *John was expected to help Mary.*

(ii) *It was expected that John would help Mary.*

Semantically, the verb *expect* – unlike *seem* – also θ-marks an external argument. This is shown by the fact that the active use of *expect* allows a referring NP in the matrix grammatical subject position when the predicate is followed by a finite clause, which is not the case with *seem*.

(iii) *John expects that George will help Mary.*

(iv) **John seems that George will help Mary.*

the *Sketch Engine* concordancer (Kilgarriff et al. 2014), we have extracted the two constructions using the CQL queries in (4) and (5).

(4) *CQL* query for extracting the raising construction

[lemma = "likely|appear|seem"] [lemma = "not"]? [lemma = "to"] [tag = "V.*"]

(5) *CQL* query for extracting the extraposition construction

[lemma = "it"] [] {0,5} [lemma = "likely|appear|seem"] [lemma = "that"] within <s/>

In plain English, the command string in (4) ensures that the *Sketch Engine* returns those constructions where the lemmas *likely*, *appear* and *seem* are immediately followed by the lemma *to*, which is in turn immediately followed by any kind of verbal expression, be it a lexical verb (e.g., *John is likely to* ***love*** *Mary*) or one of the two auxiliary verbs that can occur in infinitival constructions (e.g., *John is likely to* ***have*** *loved Mary*; *John was likely to* ***be*** *loved by Mary*). The optional condition *[lemma = "not"]?* ensures that examples with the negative marker *not* placed between the predicate and the infinitival marker, such as *John appears* ***not to*** *like Mary*, are also included. To our knowledge, this provides us with an exhaustive extraction of raising constructions from a corpus without irrelevant data. Similarly, the search string in (5) specifies that the concordancer looks only for those constructions in which the lemma *it* is followed by the sequence of *likely/seem(s)/appear(s) + that* within the same sentence. Crucially, the parameter []{0,5} ensures that up to 5 lexemes can intervene between *it* and the raising predicates, thus taking into account constructions like *It would have been likely that*, where *would*, *have* and *been* are the interveners.

2.2. The Historical Distribution in the *Brown Family* Corpus

The results of our survey of the *Brown Family*, conducted by using the method of extraction outlined in Section 2.1, are shown in Table 1.

Table 1: The diachronic distribution of the raising and extraposition constructions in the *Brown Family* corpus

		appear				seem				likely			
Year	N (words)	raising		extrapos.		raising		extrapos.		raising		extrapos.	
1931	1,002,879	135	**83%**	27	**17%**	356	**92%**	29	**8%**	95	**93%**	7	**7%**
1961	1,942,151	211	**83%**	42	**17%**	757	**91%**	71	**9%**	223	**86%**	35	**14%**
1991	1,979,237	222	**88%**	30	**12%**	641	**92%**	52	**8%**	260	**86%**	43	**14%**
2006	2,002,823	194	**91%**	20	**9%**	545	**92%**	48	**8%**	273	**88%**	38	**12%**

Two distributional facts can be observed here. First, raising constructions are significantly more frequent than the extraposition constructions in all four subcorpora – compare, for instance, the 211 (83%) instances of *appear* used in the raising construction in the 1961 corpus with the 42 instances (17%) of the same predicate appearing in the extraposition construction within the same year. This observation is in line with the idea that there is a preference for using a non-finite clause in written English in case a finite variant can also be chosen without an interpretative difference (Leech et al. 2009).

Second, the distribution of the two constructions remains generally unchanged from the diachronic perspective – a prominent change can only be observed with the predicate *appear*, in which case the frequency of the raising construction increases from 83% in the 1931 and 1961 subcorpora to 88% in the 1991 subcorpus and finally 91% in the 2006 subcorpus of all examples with the predicate. In the case of the predicate *seem*, by contrast, no such increase can be observed, as shown by the fact that the raising construction consistently accounts for 91–92% of all examples in all four subcorpora. In addition, the use of the raising construction even becomes less frequent in the case of the adjective *likely*, as shown by the drop in frequency from 93% of all examples in 1931 to 88% in 2006.

Crucially, such a diachronically unchanging distribution of the two constructions, in which the frequency of raising constructions does not increase at the expense of the semantically equivalent extraposition variants from 1931 to 2006, is inconsistent with the diachronic generalisation which states that "[i]nfinitival clauses have been gaining ground at the expense of finite clauses since the late Old English period" (Leech 2009: 204).

A concrete example of Leech et al.'s (2009) generalisation is shown in a corpus study conducted by Malá (2017). In contrast to the narrow focus of the present study, she looks at the diachronic distribution of all the three major structural patterns in which non-finite clauses appear (the examples in the parentheses are taken from Malá's (2017) own corpus findings):

(i) participial clauses functioning as modifiers in NPs (e.g., *The campaign* ***leading to the election*** *was so quiet*);

(ii) participial clauses functioning as sentential adjuncts (e.g., ***Planning a Christmas wedding****, Andrew is also preparing for high office*) and nominal (i.e., gerundial) *-ing* clauses (e.g., *Yet the UK and the USA … have blatantly pursued policies directed at* ***keeping refugees*** **out)**;

(iii) the various uses of infinitival clauses, which also subsume the embedded clauses in the raising construction (e.g., *He said legislation … is expected* ***to be introduced Monday***)

In order to observe the diachronic increase, Malá (2017: 153) uses the 1961 *Brown* and *LOB* corpora and the 1991 *Frown* and *F-LOB* corpora, which replicate the design of the former two in terms of size and text sources (cf. Mair 1997 for detailed metadata describing each individual corpus) and thus serve as comparable corpora. By analysing data automatically extracted from these corpora, Malá (2017) concludes that when groups (i)–(iii) "were [diachronically] considered as wholes, a significant increase in the frequency of non-finite clauses was noted […] [and that this] may be considered significant because there is only a 30-year difference between the Brown/Frown and LOB/F-LOB corpora." (2017: 164).[4] For instance, in the case of group (i) – participial clauses functioning as modifiers in NPs – she reports that the number of examples rose from 16,114 in 1961 to 17,264 in 1991 (2017: 157), so roughly a 7.1% increase in thirty years, with similar increase rates reported for groups (ii) and (iii).

In other words, Malá's (2017) findings show that non-finite clauses in general are becoming more and more frequently used in the 20th century, which is in line with Leech et al.'s (2009) diachronic generalisation, whereas our findings in Table 1 show that a (relatively small but syntactically distinct) subset of non-finite clauses – i.e., those appearing in the raising construction – resist this trend. The question, then, is why this is so.

2.3. The Contemporary Distribution in the *English Web Corpus 2015*

At this point, one could argue that the *Brown Family* is too small a corpus for us to draw meaningful conclusions about changes in the diachronic distribution of the two constructions, as shown by the fact that there are as few as 7 examples of the extraposition constructions with the raising adjective *likely* in the 1931 subcorpus.[5] For this reason, we have

[4] A previous diachronic study into finite vs. non-finite variation conducted by Malá (2013) similarly reports that "current academic texts in comparison with the texts written a hundred years ago display a greater tendency towards a non-finite mode of expression" (2013: 27–28) and that generally "current users of the [English] language seem to employ non-finite forms more frequently now than a hundred years ago" (2013: 28).

[5] A similar concern is shared by Leech et al. (2009), who claim that "the fact that the thirty-year period separating Brown [i.e. 1961] and Frown [i.e. 1991] corpora, or even the

conducted a follow-up survey of the distribution of the two constructions based on the more contemporary *English Web Corpus 2015*. This corpus, which is carefully annotated for part-of-speech categories and lemmatised (Jakubíček et al. 2013), contains 15.7 billion words, so it is roughly 2000 times larger in size than the *Brown Family*. In terms of content, it is webcrawled from English top-level-domain websites, which means it contains written English produced by native speakers and is therefore suitable for the study of contemporary English usage.

The results of our survey of the *English Web Corpus 2015*, also conducted by using the CQL search strings in (4) and (5), are shown in Table 2.

sixty-year period covered by B-LOB [i.e. 1931], LOB [i.e. 1961] and F-LOB [1991], are far too short to present more than a brief episode in the development of changes which usually take several centuries to complete" (2009: 204), also noting that the relatively small size of the listed corpora may pose a problem for fashioning generalisations about diachronic change (*ibid*.). Note, however, that this quote should be viewed in light of the fact that Leech et al. (2009) are also concerned with large-scale changes in the English language, such as the fact that the entire system of clausal complementation changed from Old English to Modern English.

Second, it is precisely because we share this concern that we have taken into account the distribution of raising vs. extraposition in the contemporary *English Web Corpus 2015*, which is much larger in size than the *Brown Family*. In addition, data from the *English Web Corpus 2015* extend the temporal scope of our study from the period of 1931–2006 (*Brown Family,* Table 1) to 1931–2015, so that we are able to observe a period of a total of 84 years.

Finally in relation to the issue of size, we note that the number of tokens taken into account in our survey of extraposition vs. raising in the *Brown Family* is similar to that reported in Leech et al.'s (2009) study of the predicate *want*, where the use of *want* as a catenative verb complemented by a *to*-infinitive clause increases from 325 tokens in the 1961 *Brown* corpus to 553 tokens in the 1991 *Frown* corpus (2009: 200). Crucially, this increase is taken by Leech et al. (*ibid*.) to be a significant correlate of the on-going grammaticalization process of *want* (i.e., the original modal meaning of *want* – i.e., desire – has become semantically bleached, which is shown by the fact that this verb expresses different modal meanings in different contexts, such as a suggestion (and not the original meaning of desire) in Leech et al.'s example *You don't want to look at the eclipse* (2009: 200). Note that Table 1 reports similar numerical figures – e.g., the use of raising *seem* increases from 356 instances in 1931 to 757 instances in 1961. Of course, the number of extraposition examples with *seem* increases at basically the same rate in tandem, which means that non-finite raising does not increase *at the expense* of extraposition – if it were otherwise, i.e., where only the number of the raising *seem* examples increased, then the diachronic distribution of these structures would be consistent with Leech et al.'s (2009) generalisation that non-finite clauses continue becoming more prominent in English.

Table 2: The distribution of the raising and extraposition constructions in the *English Web Corpus 2015*

		appear				seem				likely			
Year	N	raising		extrapos.		raising		extrapos.		raising		extrapos.	
2015	15.7B	857k	**85%**	149k	**15%**	2,102k	**90%**	238k	**10%**	1,075k	**89%**	127k	**11%**

The distribution of raising vs. extraposition in Table 2 is consistent with that in the *Brown Family* corpus in Table 1. What is crucial here is that the frequency of the raising construction has not increased from the older frequencies attested in the *Brown Family*. For instance, there is a 90% preference for the raising construction in the case of the predicate *seem* in 2015, which is virtually unchanged from the state of affairs in all the four subcorpora in Table 1, where there is a consistent 91–92% preference for using *seem* in the raising construction. In other words, if the raising vs. extraposition alternation conformed to the trend of the increasing use of non-finite constructions, as reported by Malá (2017), then we would expect a much higher preference for raising in the *English Web Corpus 2015*, which is not the case.

Note that many non-finite structures are derived via syntactic mechanisms that are arguably less complex than the cross-clausal syntactic movement of the raising construction. For instance, non-finite clauses functioning as modifiers in complex NPs, such as the participial clause in Malá's example *The campaign* ***leading to the election*** *was so quiet*, do not involve any kind of syntactic movement; rather, they enter the structure via the same compositional mechanism as regular adjectival phrases – specifically, the participial clause *leading to the election* functions formally as a predicate abstract denoting the property *x is leading to the election* (much like the AdjP *happy* denotes the property *x is happy*), so it attaches to the NP *The campaign* in order to narrow down its denotation to a unique individual that has the properties *x is a campaign* and *x is leading to the election*, cf. Bruening (2014) for details. Consequently, we believe that there are structural factors specific to the comparatively more complex syntax of the raising/extraposition alternation that explain why, rather exceptionally in comparison to all structures allowing speakers the choice between finite and non-finite ways of expressing information, the raising (i.e., non-finite) construction has not become more prominent at the expense of the extraposition (i.e., finite) construction diachronically. These constructions are discussed in the remainder of the paper.

3. A Synchronic Analysis of the Contemporary Distribution of Raising and Extraposition

3.1. Previous Account

Synchronic factors influencing the use of the raising construction over the extraposition one and vice versa are seldom discussed in the descriptive literature. What discussion there is often ties the choice between the two constructions to the influence of the surrounding discourse and intra-clausal information packaging. For the raising construction, Biber et al. (1999: 731–732) identify two major factors in this respect, the first being the short distance between an NP in a preceding sentence and its coreferential pronoun in the raising construction (6), and the second the heaviness of the subject NP (7).

(6) *The first thing that he thought of when he woke up was* ***Marge****.* ***She*** *wasn't likely to take a taxi to Naples.*

(7) ***The release of gases from the interior due to internal heat and chemical reaction*** *is generally thought to be the primary concern.*

For the extraposition construction, Biber et al. (1999: 733–734) discuss two factors, each of which they claim is specific to a different raising predicate. In the case of the raising verb *seem*, Biber et al. *(ibid.)* cite frequency data from their own corpus findings which show that over 98% of all extraposed constructions headed by *seem* contain an intervening *to*-phrase following the matrix predicate (8).[6]

(8) *It seemed* ***to him*** *that his home life was disintegrating all at once.*

For the raising adjective *likely*, Biber et al. (1999: 734) claim that the extraposition construction is preferred in case there is an explicit need to mark the tense or modality in the embedded *that*-clause. According to them, the modal verb *will* is the most common tense/modal marker, whereas "*would* and *might* are also used occasionally" (*ibid.*):

[6] We tentatively note that the state of affairs in the *English Web Corpus 2015* seems to contradict this empirical claim – intervening *to*-phrases occur only in about 10% of all extraposition examples with *seem*. However, a more in-depth look at the distribution of intervening phrases is left for further research, as such structures are difficult to extract from a corpus as large as the *English Web Corpus 2015* because full-fledged NPs of varying complexity (and not only simplex pronouns as in (8)) can serve as objects to the intervening prepositions (e.g., *It seemed to* ***my childhood friend*** *that the old days were over*).

(9) a. *It is likely that North Korea **will** channel investment to areas that can be contained.*

b. *It is likely that a moose to the right **might** be discerned easier than if it appears on the left side.*

In Section 3.3, we will show that the synchronic state of affairs goes beyond information structuring and tense marking and introduce novel structural factors that constrain the choice between the two constructions.

3.2. The Set-Up of the Synchronic Analysis

The synchronic analysis of the extraposition and raising constructions was done on the basis of two randomized samples, one containing 200 examples of the raising construction and the other 200 examples of the extraposition construction, extracted from the *English Web Corpus 2015*. The same three predicates as in Section 2 were taken into account – *seem*, *appear*, and *likely*. The samples were manually annotated for recurring structural phenomena, whose prominence was then checked against the entire corpus.

The analysis has produced two results. First, we have managed to identify a wholly novel syntactic factor, discussed in Section 3.3.1, that partially explains why the raising construction has diachronically been much more prominent than the extraposition construction. Second, we show that there are two prominent contexts – one structural (Section 3.3.2) and the other pragmatic (Section 3.3.3.) – for which the extraposition construction is better suited than the raising variant.

3.3. The Identified Factors

3.3.1. The Use of Relative Pronouns Favours Raising

The most frequently occurring structural phenomenon in our raising sample (28 out of 200 examples) is the use of a relative pronoun at the left clausal edge of the construction, as seen in the following examples taken from the *English Web Corpus 2015*:

(10) *Individuals who are likely to require personal care under the state plane, or home* […]

(11) *I have limited how much I eat* [,] *which seems to be a small step in the right direction.*

(12) *And that his involvement with what appears to be a Move to make the NEA push Obama's agenda is being done pro-bono.*

There are two broader syntactic patterns here. The first is shown by examples (10) and (11), in which the relative pronouns *which* and *who* function as the logical subjects of embedded raising constructions and are anaphorically related to expressions within the same sentence. In example (10), the raising construction is a postmodifier in the NP headed by the noun *individuals*, with which the pronoun is coreferential, so it is used as an adjectival relative clause. In example (11), which is in terms of anaphoric relations a sentential relative clause, the pronoun *which* refers back to the entire proposition expressed by the matrix clause. The second pattern is shown by example (12), where the raising construction functions as a nominal relative clause that is the syntactic object of the preposition *with*.

By contrast, there are no such constructions in the extraposition sample. Intuitively, this is congruent with the idea that the constructed example in (13b), in which the nominal relative clause is an extraposition construction, seems less natural than the version in (13a), where the relative clause is a raising construction.[7]

(13) a. *We must take what appears to be a detour.*

b. ??*We must take what it appears is a detour.*

In the entire *English Web Corpus 2015*, however, one can in fact find examples like (13b), as shown by sentences (14) – (16), which are taken from the corpus.[8]

[7] Note that sentence (13a) is a raising construction because *what* starts off as an underlying constituent of a small clause SC that also contains the NP *a detour*, as in the following simplified derivation:

(i) *We must take what*$_i$ *appears* [$_{TP}$ *to be* [$_{SC}$ ~~*what*~~$_i$ *a detour*]].

The relative pronoun *what* enters the syntactic derivation alongside the NP *a detour* because a predication relation is established between them, whereas the matrix verb *appears* and *what* are not involved in any semantic relation (it will be shown in Figure 2 that a *wh*-pronoun like *what* moves to the matrix subject position for case-checking reasons, which are non-semantic in nature). Or to give a simpler example without relative pronouns, the grammatical subject *John* in sentence (ii) starts off within the same small clause as the AdjP *smart* because *John* is the logical argument of the property *x is smart* denoted by the AdjP.

(ii) *John*$_i$ *appears* [$_{TP}$ *to be* [$_{SC}$ ~~*John*~~$_i$ *smart*]].

[8] An anonymous reviewer suggests that a clause like *it would seem* in sentence (14) is syntactically a juxtaposed constituent, which implies that it functions as a disjunctive

(14) *I became exceedingly unhappy at that mysterious non-intercourse which it would seem had been rendered eternal between the different planets* […]

(15) *My grandfather* (sic) *who it appears was born out of wed lock was born Donald Harry Templar* […]

(16) *He is doing the wrong thing for what it seems are the right reasons.*

In terms of grammatical function, sentences (14) and (15) contain embedded extraposition constructions used as adjectival relative clauses (i.e., as postmodifiers in NPs), while sentence (16) uses the construction as a nominal clause functioning as the object of the preposition *for*. These examples thus show that there are extraposition constructions in actual use which play the same embedded syntactic roles (and concomitantly pattern with the same three relative pronouns) as the raising structures in (10)–(12).[9] Note that the extraposition pattern in (14)–(16) is also semantically equivalent to the raising pattern in (10)–(12).

Although attested, such extraposition examples turn out to be extremely infrequent in comparison to the raising variant. Table 3 shows the frequency of their occurrences in the corpus for all three predicates, contrasted with that of the raising construction in the same grammatical role.

---

modifier that can be omitted without a change in meaning. This is not so – the embedded *it (would) seem(s)* and *it appears* clauses in sentences (14)–(16) do in fact make a crucial contribution to the semantics of the sentences, as they introduce an epistemic modal interpretation (corresponding to a strong probability) to their embedded clauses. This is shown by the fact that e.g. sentence (16) can truthfully be uttered in a situation in which whatever is denoted by *what* happens not to correspond to "the right reasons". This would not be the case if *it seems* were omitted. Crucially, the embedded clauses from which the pronouns are extracted syntactically function as the complements of the superordinate *it seems* and *it appears* clauses (as is the case in Figure 1 below), since only under such a configuration can a modal meaning be ascribed to the embedded clauses (cf. Heim and Kratzer 1998: Chapter 12).

[9] We only take into account extraposition and raising constructions in which the relative pronoun logically corresponds to either the underlying subject of an active clause or the underlying object of a passive clause. Constructions in which the pronoun corresponds to the logical object of an active clause, as in (i), are not discussed here, since such structures do not have a corresponding raising variant, as is shown by the ungrammaticality of (ii).

(i) *My friend who it seems John likes ~~who~~ is in love with me.*

(ii) **My friend who seems John to like ~~who~~ is in love with me.*

Table 3: The distribution of extraposition and raising constructions with relative *wh*-pronouns in the *English Web Corpus 2015*

extraposition		**raising**	
all examples	with RP as subject	all examples	with RP as subject
514,814	522 (**0.1%**)	4,034,844	188,462 (**4.7%**)

Extraposition constructions with relative pronouns as subjects like in (14) – (16) account for only about 0.1% of all the 514,814 extraposition clauses in the corpus. In comparison to the raising constructions (4.6% of all examples), this is an extremely low frequency of occurrence that highlights the idea that extraposition constructions are dispreferred in the syntactic role of relative clauses.

We now provide a comparative analysis of the syntactic derivations that underline the raising and extraposition constructions with relative pronouns as grammatical subjects. We show that the extraposition construction involves a greater number of derivational steps, both related to agreement and syntactic movement, than the raising construction, which is why it is the cognitively most complex structure among the two.

We start by discussing the derivational steps of the extraposition relative clause in sentence (15), *who it appears was born* (*out of wedlock*). The full derivation is shown in Figure 1.

There are two derivational operations in Figure 1 that are crucial for our purposes. These are Agree and Move, both of which are couched in syntactic assumptions that underlie Chomsky's (2001) *Derivation by Phase* framework. Move is the syntactic operation that selects a structural constituent and re-merges it in a higher position within the structure, leaving an unpronounced copy (shown by strikethrough) in the extraction site.

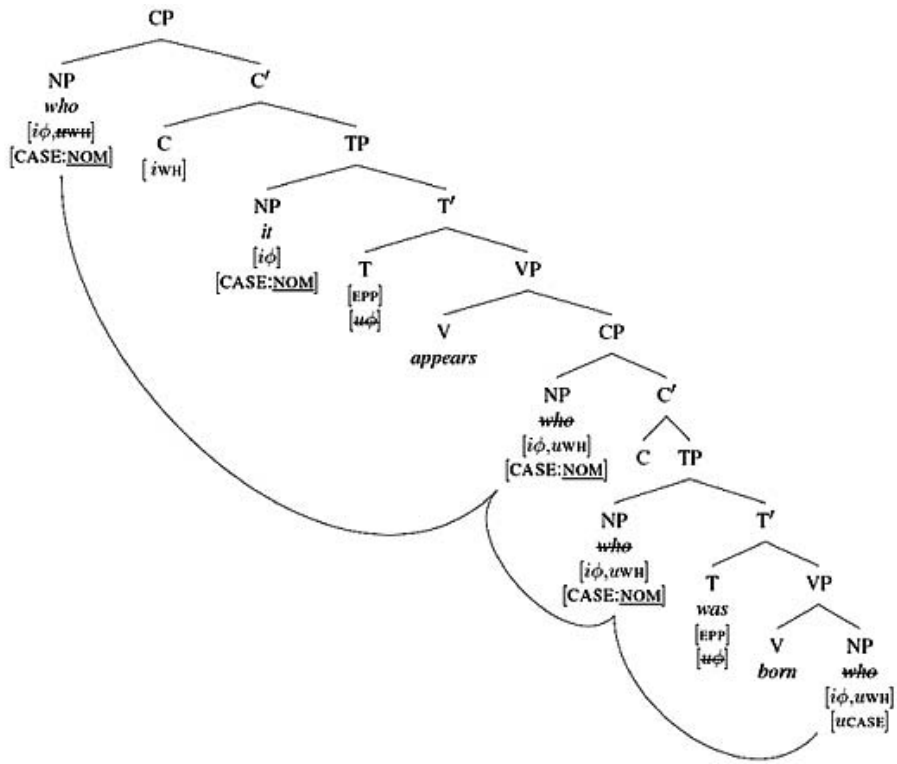

Figure 1: The syntactic derivation of the extraposed relative clause in sentence (15)

In this framework, Move is licenced by the Agree operation, which Chomsky (2001: 3–4) defines in terms of feature agreement between two constituents that are in an active Probe-Goal relation, where an active Probe is a structural node bearing semantically uninterpretable features while a Goal is a node that is *c-commanded* by a Probe and bears matching interpretable features.[10] In Figure 1, the head of the embedded tense projection T is a probe

[10] The syntactic relation *c-command* is defined as follows:

(i) A constituent X c-commands its sister constituent Y and any constituent Z which is contained within Y. (Radford 2004: 91)

From definition (i) it follows that e.g. the embedded constituent T in Figure 1 c-commands its complement VP (which is its sister) and both V *born* and NP ~~*who*~~, which are the constituents contained in VP.

with uninterpretable *φ*-features (glossed as [*uφ*]), which finds its goal with matching features [*iφ*] in the internal argument NP – i.e., the relative pronoun *who*, which complements the participle *born*. At this point, two syntactic operations take place – on the one hand, the uninterpretable features on T are deleted under agreement, which is again shown by strikethrough, while on the other, this self-same operation induces the so-called EPP property, which in Chomsky's (2001) system is employed to displace *who* from its initial position as the complement of V to the Spec(ifier) position of TP, this being the first movement in the derivation. Crucially and as a by-product of *φ*-feature agreement between T and NP, the uninterpretable case feature [*u*CASE] on NP is valued as nominative (Chomsky 2001: 6).

Because *φ*-feature agreement takes place between the lower T and NP, the embedded clause is finite. In Chomsky's (2001) system, this means that the clause (or, more technically, the embedded CP projection) constitutes a "strong phase" of derivation, which – in simplified terms – means that all the derivational relations between the embedded constituents have to be finalized before the derivation of the next strong phase – i.e., the matrix clause in Figure 1 – takes place, the idea being that a division of the derivation into separate phases reduces the "cognitive" computational burden of structure building (Chomsky 2001: 11–12). As a crucial consequence of this reduction, all constituents within the lower strong phase become inaccessible to syntactic operations when the derivation of the higher phase starts taking place. To capture this, Chomsky (2001: 13) proposes the Phase Impenetrability Condition, which is given in (17) and is defined for any strong phase projection HP.

(17) The [c-command] domain of H is not accessible to operations outside HP; only H and its *edge* are accessible to such operations.

In Figure 1, the embedded CP is a strong phase, while the accessible *edge* by definition (17) corresponds to the Spec position of CP. Since the relative pronoun *who* needs to move outside the embedded clause in order to get into a position where it can check its [*u*WH] feature against a matching goal, Move displaces *who* from the Spec position of the embedded TP into the Spec position of the embedded CP (which is a strong phase), where it becomes – again by definition (17) – accessible to operations in the next phase. This is the second step of movement that takes place in the derivation.

What then follows in the matrix clause is that another Agree operation takes place, which is now triggered by the uninterpretable *φ*-features in the

matrix T. Since the case feature on the relative pronoun was already valued in the lower clause, *who* cannot step into an agreement relation with the higher T, so the expletive *it* is merged in the matrix Spec, TP position where it agrees with and deletes the uninterpretable features on T.

Finally, the relative pronoun *who*, which is still a syntactically active constituent due to the [*u*WH] feature, moves from the Spec position of the embedded CP to the Spec of the matrix CP, where it can finally delete its uninterpretable [*u*WH] feature because it is in a position where it c-commands the matrix C head, which has the matching feature [*i*WH]. In short, the derivation of the extraposition clause *who it appears was born* in Figure 1 involves three applications of Move and two instances of Agree.

By contrast, the raising variant of the extraposed relative clause in (15), i.e., *who appears to have been born*, involves – in a system like Chomsky's *Derivation by Phase* (2001) – far fewer operations. This is shown in Figure 2.

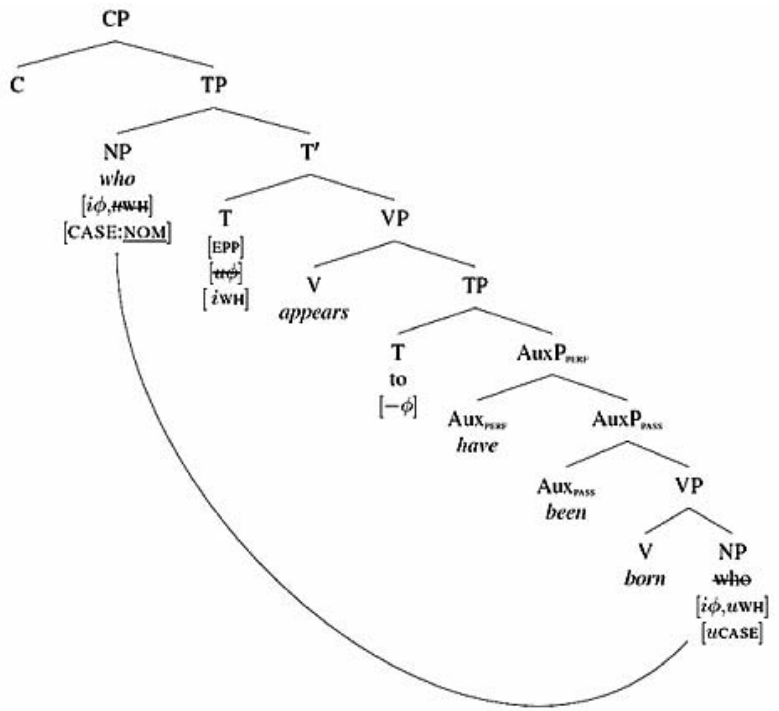

Figure 2: The derivation of the raising relative clause who appears to have been born

Here, the head of the TP of the infinitival embedded clause is structurally impoverished because it is "*φ*-incomplete" (Chomsky 2001: 8) – i.e., it lacks a set of uninterpretable *φ*-features, which means that *who* does not check and value its [*u*CASE] feature against the embedded T node, and consequently does not move to the Spec position of the embedded TP, in contrast to what happens in the embedded clause in Figure 1. (Recall that this explains the ungrammaticality of sentence (3b), in which the grammatical subject stays put in the embedded position.) In addition, because a *φ*-complete T head is a prerequisite for the projection of CP in Chomsky's (2001) system, the embedded clause in Figure 2 lacks the CP layer, which means that it does not constitute a strong phase of derivation, again in contrast to the finite embedded clause in Figure 1. As a consequence, *who* enters into an Agree relation with the matrix T, and moves directly into its Spec position, in which its structural case feature gets valued as nominative.

The [*u*WH] feature in *who* gets checked in the Spec, TP position as well. This is so because an additional application of the Move operation, which would displace *who* from the Spec of the matrix TP to the Spec of the matrix CP (i.e., the final position of *who* in Figure 1), would violate the *Vacuous Movement Hypothesis*, which was originally proposed by George (1980) and Chomsky (1986) and which states that "vacuous movement, i.e., a movement operation whose effect cannot be observed, should not be allowed" (Ishii 2004: 187). In such a configuration, where *who* remains in the Spec of the matrix TP, the [*i*WH] feature is transferred from C to T under Sakamoto's (2012) proposal of Feature Inheritance, which is an operation that transfers the features of a clausal edge, i.e., [*i*WH] in our case, to a position which "enables feature valuation to be implemented in one derivational position" (Sakamoto 2012: 323). That is, *who* checks both its [*u*CASE] feature and its [*u*WH] feature in one fell swoop against T. According to Sakamoto (*ibid*.), such a structural assumption is computationally more optimal than a configuration which posits that the [*i*WH] feature is present on the separate C head, as this would involve vacuous movement.

In short, the derivation of the raising relative clause in Figure 2 involves only one instance of the Move operation and one instance of Agree, in contrast with the extraposed clause in Figure 1, which involves three instances of Move and two instances of Agree. The fact that Move is in syntactic minimalism taken to be a complex operation that needs to be licenced by feature-agreement means that its application is a cognitively

costly procedure (hence Chomsky's (2001) proposal that syntactic derivation takes places in separate phases). For the distinction between raising and extraposed constructions with relative pronouns as subjects, this means that the former will serve as a speaker's preferred means of encoding information due to the comparatively few instances of derivational steps involved.

It is also worth noting that in Orešnik's (2007) Natural Syntax framework, which is a deductive linguistic theory that aims to establish a scale of naturalness (or markedness) between a set of closely related competing constructions, the extraposition construction in Figure 1 would be seen as less "natural"/more marked (though still grammatically acceptable) than the raising construction in Figure 2 precisely because its more complex derivation, involving several steps of movement, is not in accordance with what Orešnik (2007) calls the Principle of Least Effort, which states that among two semantically equivalent but structurally distinct constructions, the one that is the most natural is the one that "is cognitively simple (for the speaker) [i.e.,] easy to produce, easy to retrieve from memory, etc." (2007: 35).

In addition to its complex derivation, the low frequency of extraposition in Table 3 is due to a formal syntactic constraint called the *that*-trace effect, originally observed by Perlmutter (1971). This constraint refers to the fact that subject extraction is not allowed if the subordinator *that* is present in the clause in which the subject originates (18). Note that this effect is structurally asymmetric in that it does not hold for object NPs, which can be extracted in the presence of *that* in the embedded clause (19).[11]

(18) a. *Who do you think ~~who~~ met Sue?*
b. **Who do you think that ~~who~~ met Sue?*

(19) a. *Whom do you think John met ~~whom~~?*
b. *Whom do you think that John met ~~whom~~?*

(Examples from Pesetsky 2017)

[11] The *that*-trace effect is only one instance of a larger set of syntactic asymmetries that hold between subjects and objects. Another prominent example relates to the fact that extraction out of complex NPs is only allowed if they function as structural objects (ii). For a contemporary formal analysis of these facts, we refer the reader to Bošković (2018).

(i) **Who did friends of who hire Mary?*
cf. *Friends of John hired Mary.*

(ii) *Who does Mary sell pictures of who?*
cf. *Mary sells pictures of John.*

The extraposition construction is also constrained by the *that*-trace effect. This is shown by the fact that the sentences in (14) – (16) all become ungrammatical if *that* is added to the embedded extraposed clause:

(20) **I became exceedingly unhappy at that mysterious non-intercourse which it would seem* ***that*** *had been rendered eternal between the different planets* [...]

(21) **My grandfather who it appears* ***that*** *was born out of wed lock was born Donald Harry Templar* [...]

(22) **He is doing the wrong thing for what it seems* ***that*** *are the right reasons.*

In contrast to the extraposition construction, the raising variant is not constrained by the *that*-trace effect, given that the subordinator *that* does not pattern with infinitival clauses in any case. In addition, the raising construction also involves cross-clausal movement even outside the domain of *wh*-clauses, given the fact that the embedded clause is non-finite and therefore cannot host the embedded subject, as shown by the derivation in Figure 2.

We therefore conclude by proposing that speakers prefer using the raising construction in case of *wh*-subject extraction precisely because of this fact; that is, the raising construction involves cross-clausal subject movement by default, whereas applying *wh*-movement to the extraposition construction, where the embedded subject stays put in case it is not realised as a *wh*-pronoun, is a more costly way of encoding information by comparison, especially given the multiple movement analysis of Figure 1.

3.3.2. The Use of Negation is Syntactically Constrained in the Raising Construction

In this Section, we show how the placement of the negative marker *not*, as well as its contracted affixal forms, is distributed across the extraposition and raising constructions in the *English Web Corpus 2015*. The four possible patterns are given in (23) – (26).

(23) The extraposition construction with matrix negation

a. [*I*]*t does not appear that indefinite back-up power is available* [...]

b. *But now it does not seem likely that I will be able to complete either one* [...]

(24) The extraposition construction with embedded negation

a. *And it seems that the Congress wasn't going to support it either*.

b. *It appears that this driver of ours is not an official tour guide* […]

(25) The raising construction with matrix negation

a. [*T*]*he defender doesn't seem to know when silence works best* […]

b. *This conclusion does not appear to be supported by the record*.

(26) The raising construction with embedded negation

a. *The agents appeared not to believe her and threatened to imprison her* […]

b. *The Android market is likely to not be included in favour of the aforementioned* […]

These are the frequencies of the patterns:

Table 4: The distribution of negation in the matrix and embedded clauses[12]

	Extraposition	**Raising**
All examples	514,814	4,034,844
All examples with NEG	30,753 (**6%**)	297,809 (**7.4%**)
Matrix NEG	7,235 (**1.5%**)	272,931 (**6.8%**)
Embedded NEG	23,518 (**4.5%**)	24,878 (**0.6%**)[1]

The negative marker is inversely distributed among the matrix and embedded clauses of the two constructions. On the one hand, the extraposition construction favours the use of the negative marker in the embedded clause (4.5% of all instances of the construction in the corpus), as in the examples under (24). On the other hand, the raising construction more frequently patterns with matrix negation (in 6.8% of all instances of the construction in the corpus), as in the examples under (25).

Crucially, the pattern in (26) – with embedded negation in the raising construction – is by far the least frequent. From a formal theoretical perspective, such a low frequency is to be expected. As shown by the

[12] Note, in addition, that the pattern in (26) comes in two structural flavours, since the negative marker can either precede the infinitival marker *to*, as in the (a) example, or follow it, as in the (b) example. The latter sequence of markers is much less frequent, and constitutes approximately a fifth of all raising examples with embedded negation.

syntactic derivation in Figure 2 in Section 3.3.1., the primary syntactic characteristic of infinitival clauses is their impoverished structure in comparison to finite clauses. Wurmbrand (2001; 2014) builds upon this observation by arguing that, cross-linguistically, infinitival clauses are not uniform in the degree of the syntactic structure they contain. The types of infinitives that have the least structure, which she dubs as *lexically restructuring infinitives*, for instance, do not allow the use of embedded negation at all. A case in point is the German predicate *erlauben* ("allow"), where negation can only take matrix scope (in contrast to English *allow*, which is thus a *non-restructuring infinitive*), as shown by the fact that the second translation, with negation embedded under *allow*, is not permissible.

(27) *Weil dem Hans der Spinat nicht zu essen erlaubt wurde*.
because the Hans.DAT the spinach.NOM not to eat allowed was
"since Hans was not allowed to eat the spinach."
*"since Hans was allowed not to eat the spinach."
(Example from Wurmbrand 2001: 118)

While the use of negation in infinitival clauses is never formally banned in English, it does show varying degrees of salience in different syntactic contexts, which is in line with Wurmbrand's claim that not all infinitives are structurally uniform in their deficiency. For instance, even less frequent than the raising pattern in (26) is the use of embedded negation with the catenative verb *want*. In the *English Web Corpus 2015*, there are only 630 examples of embedded negation with *want* followed by an infinitival clause, as in (28), whereas matrix-clause negation with *want*, as in (29), is attested in (at least) 618,888 examples. Hence, the striking infrequency of using embedded negation with *want* is reminiscent of Wurmbrand's *lexically restructuring predicates*, which disallow embedded negation altogether; indeed, it seems that negation in (28) is licenced by a special pragmatic factor – concretely, the placement of emphasis on what the referent of *she* (that is, Virginia Woolf) wishes not to be the case.

(28) *Woolf didn't want sympathy; she* ***wanted not to*** *be silenced, and to prove to Eliot, and to us, that vulnerability has its own kind of genius*.

(29) *Mr. Taylor* ***didn't want to*** *get too gloomy though.*

Like the catenative use of *want*, English raising constructions are in terms of interpretation a *NEG*-raising environment (Horn 1989). This means that the placement of negation is only a matter of the surface syntactic representation, as shown by the fact that a raising sentence like *John doesn't seem to like Mary*, with matrix negation, is interpretatively equivalent to *John seems not to like Mary*, with embedded negation (cf. Postal and Collins 2012).[13] Therefore, English speakers are presented with a non-semantic paradigmatic choice of placing the negative marker either in the matrix or embedded clause when using the raising construction.

Coupled with this fact, we believe, is that the striking infrequency of embedded negation is a reflex of the broader generalisation pertaining to the lesser amount of structure that infinitives have cross-linguistically. In other words, a prediction of the generalisation is that speakers will place the negative marker in the clause that has more structure, which explains why the matrix placement of negation is vastly more frequent in Table 4, even though the use of embedded negation is not outrightly banned from the formal perspective. By contrast, the extraposition construction is made up of two finite (i.e., structurally non-impoverished) clauses, which is reflected in the smaller gap between the frequency of extraposition with matrix negation (23) and the frequency of extraposition with embedded negation (24). In other words, the extraposition construction provides speakers with a relatively unconstrained environment for placing negation, whereas the raising construction presents a more limited option, in that only the matrix clause is finite.

[13] In terms of formal logic, negation is invariantly interpreted in the embedded position in *NEG*-raising environments, as shown by the fact that in the *NEG*-raised variant *John doesn't seem to like Mary*, the message is about *whom John doesn't like* and not about *what doesn't seem*. In addition, certain predicates allow *NEG*-raising only in certain syntactic configurations; for instance, *wish* only allows it if it is complemented by an infinitival clause (Quirk et al. 1985: 1034):

(i) a. *I wish not to worry you.*
b. *I don't wish to worry you.*
(ii) a. *I wish that my friend hadn't died.*
b. * *I don't wish that my friend had died.*

3.3.3. The Use of Modal Verbs in the Matrix Clause Favours Extraposition

In this Section, we look at how modal verbs are used in the two constructions. Specifically, we determine whether there is a prominent distinction in the frequency at which the two constructions contain modal verbs in the matrix clause, as in patterns (30) and (31), and discuss the consequent implications of this distributional distinction.

(30) Modal verbs in the matrix clause of the extraposition construction

a. *It would appear that the two royal princesses were accustomed to visit the same fountain in the early morn* [...]

b. [*I*]*t might seem that I was somehow pre-empting the Maori claim.*

(31) Modal verbs in the matrix clause of the raising construction

a. *Such an exercise would seem to be a great way to introduce media literacy concepts into the biology classroom.*

b. [*I*]*t is not hard to see that this would be likely to result in monetary instability.*

These are the frequencies of the two patterns:

Table 5: The distribution of modals in the matrix clauses of the extraposition and raising constructions

	Extraposition	**Raising**
All examples	514,814	4,034,844
Modal in matrix clause	42,980 (**8.3%**)	83,655 (**2%**)

The extraposition construction (30) uses modal verbs roughly four times more frequently than the raising construction (31) – i.e., in almost a tenth of all extraposition examples in the corpus.

Let's first turn to the fact that the most common modal verb in the matrix clause in both constructions is *would*, which occurs in almost two thirds of all examples in the corpus. Crucially, this modal verb, which appears in its preterite form, is in these sentences to a large extent semantically bleached given that epistemic modality is already lexically inherent to the raising predicates themselves (cf. the fact that sentence (30a) retains its modal interpretation even if *would* is dropped in the matrix

clause). Such modal verbs are rather used primarily as emphatic expressions highlighting the tentativeness that pragmatically arises by way of the epistemic possibility meaning independently denoted by the raising expression. Note that, in the functionalist literature, the epistemic predicates *seem*, *appear* and *likely* are analysed as *lexical hedges* (cf. Schmied 2008: 91 for additional examples), which are lexemes whose primary function is "to make things more or less fuzzy" (Lakoff 1972: 194) – that is, they are used to convey that the speaker is not fully committed to his or her assertion, and the additional modal is a marker of that.

Given this background, we believe that the different surface syntactic representations of the two constructions explain the distribution of modal verbs in Table 5, to the effect that the clausal structure of the extraposition construction lends itself particularly well for expressing tentativeness. To see this, compare the surface structure of the (non-corpus) extraposition sentence in (32a) with that of its raising alternative in (32b).

(32) a. *It would appear that John loves Mary.*
b. *John would appear to love Mary.*

In the surface representation of the extraposition construction (32a), the constituents of the matrix clause are to a large extent devoid of semantic content, given that the grammatical subject is a dummy pronoun and *would* is used as a weak modal that emphasises the speaker's tentativeness about the assertion in the embedded clause.[14] This is not so in the raising construction, since the underlying subject of the embedded clause raises to become the grammatical subject of the matrix clause. Consequently, the entire first clause in sentence (32a) – but not in (32b) – functions as a compositional hedge, and the greater frequency of modal

[14] Related to this observation is the fact that extraposition constructions tend to resist syntactic environments where they are immediately preceded by a functional word. For instance, there are fewer than 6500 examples (which is about 1.2% of all extraposition constructions) in the corpus in which the construction is immediately preceded by the subordinator *that*, as in sentence (i), which feels stilted because of the cluster of functional items in bold. By contrast, the raising construction is preceded by *that*, as in (ii), in about 5% of all raising examples.

(i) [I]t is great **that it** seems **that** playing the first two games is not required to know the story of The Witcher 3.
(ii) I noticed that the project seems to be closed to new members of the online community.

verbs in the matrix clause of extraposition constructions reflects precisely this fact. That is, the extraposition construction is a pragmatic device that is even more suited for expressing tentativeness than the otherwise interpretatively equivalent raising construction, as it keeps all the pragmatic (i.e., propositionally non-essential) content separate from the constituents that make up the actual assertion of the sentence.

4. Conclusion

In this paper, we have presented a diachronic and synchronic corpus-based analysis of two semantically equivalent constructions – the raising construction, in which an epistemic predicate is followed by a non-finite clause as its sole semantic argument (e.g., *John seems to love Mary*), and the extraposition construction, in which the same predicate introduces a finite clause (e.g., *It seems that John loves Mary*).

We have first established two diachronic facts on the basis of the *Brown Family* corpus, which provides data from 1931 to 2006, and the contemporary *English Web Corpus 2015*. First, the raising construction is much more frequently used than the extraposition construction in the observed time span from the beginning of the 20th century to 2015. Second, the distribution of the frequencies of the two constructions remains more or less consistent within this period. We have shown that such a diachronically consistent distribution of the two constructions stands in contrast to the observation made in the descriptive literature that non-finite forms of expressions have generally become more prominent in contemporary written English, while their finite variants have become increasingly less used.

The synchronic analysis has tied the exceptional diachronic distribution of the raising and extraposition constructions to distinct structural features associated with each type of construction. We have identified a wholly novel factor which partially explains why the raising construction is much more prominent in actual usage: the raising construction is frequently used as a relative clause in which the moved constituent is realised as a *wh*-pronoun, whereas the extraposition variant largely resists such a syntactic role.

Subsequently, we have identified two structural contexts for which the extraposition construction is better suited than its raising variant; this, in turn, explains the diachronic stability of its frequency of use, which is

exceptional in comparison to other structures allowing for finite/non-finite variation. First, we have shown that the extraposition construction patterns frequently both with matrix and embedded negation, whereas the raising variant only seldom uses embedded negation. We have argued that this distinction follows from the cross-linguistic generalisation which states that non-finite clauses are structurally impoverished and are thus less suitable for hosting functional categories than finite clauses are. Second, our analysis has shown that the extraposition construction uses modal verbs in the matrix clause significantly more frequently than the raising construction, which we have argued is tied to the idea that the clausal composition of the extraposition construction is structurally more suited for expressing tentativeness.

References

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999) *Longman Grammar of Spoken and Written English*, Pearson Education Limited, Harlow.

Bošković, Željko (2018) "On Movement out of Moved Elements, Labels, and Phases", *Linguistic Inquiry*, 49, 2, 247–282.

Bruening, Benjamin (2014) "Word formation is syntactic: Adjectival passives in English", *Natural Language & Linguistic Theory*, 32, 2, 363–422.

Chomsky, Noam (1986) *Barriers*, MIT Press, Cambridge, MA.

Chomsky, Noam (2001) "Derivation by Phase", *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52.

Collins, Chris, Paul M. Postal (2012) *Classical NEG Raising*, MIT Press, Cambridge, MA.

George, Leland Maurice (1980) *Analogical generalization in natural language syntax*, doctoral dissertation, Massachusetts Institute of Technology.

Heim, Irene, Angelika Kratzer (1998) *Semantics in Generative Grammar*, Wiley-Blackwell, Hoboken, New Jersey.

Horn, Laurence (1989) *A Natural History of Negation*, University of Chicago Press, Chicago.

Ishii, Toru (2004) "The Phase Impenetrability Condition, the Vacuous Movement Hypothesis, and *that-t* effects", *Lingua*, 114, 2, 183–215.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovár, Pavel Rychlý, Vít Suchomel (2013) "The TenTen Corpus Family", 7th *International corpus Linguistics Conference CL 2013*, 125–127.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovár, Jan Michelfeit, Pavel Rychlý, Vít Suchomel, (2014), "The Sketch Engine: ten years on", *Lexicography*, 1, 1, 7–36.

Lakoff, George P. (1972) "Hedges: A Study of Meaning Criteria and the Logic of Fuzzy Concepts", *Chicago Linguistic Society Papers* 8, 183–228.

Leech, Geoffrey, Marianne Hundt, Christian Mair, Nicholas Smith (2009) *Change in Contemporary English: A Grammatical Study*, Cambridge University Press, Cambridge.

Mair, Christian (1997) "Parallel corpora: A real-time approach to language change in progress", *Corpus-Based Studies in English: Papers from the Seventeenth International Conference on English-Language Research Based on Computerized Corpora*, ed. Magnus Ljung, 195–209.

Malá, Marcela (2013) "Notes on Norms and Usage of Finite/Non-Finite Predication in Written English", *Brno Studies in English*, 39, 1, 27–40.

Malá, Marcela (2017) "A Corpus-Based Diachronic Study of a Change in the Use of Non-Finite Clauses in Written English", *Prague Journal of English Studies*, 8, 1, 151–166.

Orešnik, Janez (2007) "Natural syntax: English interrogative main clauses", *Linguistica*, 47, 1, 35–48.

Perlmutter, David M. (1971) *Deep and surface structure constraints in syntax*, Holt, Rinehart and Winston, New York.

Pesetsky, David (2017) "Complementizer-Trace Effects", *The Wiley Blackwell Companion to Syntax, Second Edition*, https://doi.org/10.1002/9781118358733.wbsyncom108.

Pesetsky, David, Esther Torrego (2007) "The Syntax of Valuation and the Interpretability of Features", *Phrasal and Clausal Architecture*, eds. Simin Karimi, Vida Samiian and Wendy K. Wilkins, 262–294.

Postal, Paul M. (1974) *On Raising: One Rule of English Grammar and Its Theoretical Implications,* MIT Press, Cambridge MA.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*, Longman, London.

Radford, Andrew (2004) *Minimalist Syntax: Exploring the Structures of English*, Cambridge University Press, Cambridge.

Sakamoto, Akihiko (2012) "Feature Inheritance and Vacuous Movement", *English Linguistics*, 29, 2, 316–343.

Schmied, Josef (2008) "Hedges in specialised vs. popular academic interaction: A case study of medical texts, *Discourse and Interaction*, 1, 2, 85–98.

Wurmbrand, Susi (2001) *Infinitives: Restructuring and clause structure*, Mouton de Gruyter, Berlin.

Wurmbrand, Susi (2014) "Tense and Aspect in English Infinitives", *Linguistic Inquiry*, 45, 3, 403–447.

SAŽETAK

Jakob Lenardič i Gašper Ilc

ENGLESKI UZDIŽUĆI PREDIKATI I (NE)FINITNE REČENICE: DIJAKRONIJSKA I SINKRONIJSKA PERSPEKTIVA

U ovome radu predstavljamo dijakronijsku i sinkronijsku analizu konstrukcija s uzdižućim predikatima (engl. *raising predicates*) i konstrukcija s ekstrapozicijom u povijesnoj bazi *Brown Corpus* i suvremenijoj *English Web Corpus 2015*. Počinjemo utvrđivanjem dviju dijakronijskih činjenica: 1. uzdižuće (engl. *raising*) konstrukcije koriste se puno češće od semantički ekvivalentnih inačica s ekstrapozicijom, i 2. distribucija uzdižućih konstrukcija i ekstrapozicije – prilično iznimno u usporedbi s drugim strukturama koje dopuštaju varijaciju finitno/nefinitno – dijakronijski je dosljedna od početka 20. stoljeća sve do 2015. godine. Ovu jedinstvenu dijakronijsku distribuciju potom dopunjavamo analizom suvremenijega korpusa koja pokazuje da izborom između ovih dviju semantički ekvivalentnih konstrukcija upravljaju različiti strukturalni čimbenici svojstveni svakoj od njih. Konkretno, pokazujemo da se uzdižuća konstrukcija često koristi kao relativna rečenica, dok se inačica s ekstrapozicijom u toj sintaktičkoj ulozi rijetko javlja. Nasuprot tomu pokazujemo da se istaknut čimbenik u korist ekstrapozicije odnosi na oznaku niječnosti koja se podjednako često smješta i u glavnu i u uklopljenu rečenicu u slučaju konstrukcije s ekstrapozicijom, nasuprot uzdižuće inčice u kojoj se oznaka niječnosti javlja gotovo isključivo u glavnoj rečenici. Konačno, pokazujemo da konstrukcije s ekstrapozicijom češće sadrže modalne glagole u glavnoj rečenici nego li uzdižuće inačice te tu opservaciju povezujemo s idejom da je rečenična struktura konstrukcija s ekstrapozicijom strukturalno prikladnija za izražavanje nesigurnosti/uvjetnosti.

Ključne riječi: *nefinitne rečenice; ekstrapozicija; uzdizanje; korpusna lingvistika; formalna lingvistika; dijakronija*