

Complex subjects diachronically

*Erwin R. Komen
October 20th, 2010
(version #1)*

1. Introduction

SIL's linguistics discussion list recently talked about the possibility to have complex subjects in first position or not. Allan Johnson commented on the preference of (1b) over (1a) by stating:

"...it seems like "it" is there just because the grammar requires something to be there in that clause-initial position, even if only to sit in as a dummy subject because the real one doesn't fit quite comfortably enough."

- (1) a. #_[NP-SBJ] [_{CP} That Bubba shot the jukebox]] became clear.
b. _[NP-SBJ] It] became clear [_{CP} that Bubba shot the jukebox].

The comments by Allan Johnson made me wonder what has happened in English diachronically with complex subjects—those that contain any kind of CP. It could be relevant to the development of the cleft construction in English to know whether first-position subjects, or, generalizing, preverbal subjects, have changed in the way they accept complexity. If complex subjects are found less in preverbal position diachronically, then that would confirm the promotion of dummy subjects diachronically, and that in turn would offer some explanation to the rise of the it-clefts.

Armed with these ideas I have undertaken done some research to see how complex subjects as a whole fare diachronically, whether the kind of complex subjects changes over time, and what happens to the acceptability of preverbal versus postverbal complex subjects.

2. Complex subjects over time

It is not clear what changes are to be expected in the availability of complex subjects over time. Old English could be thought of as a language that is less inclined to have complexity, which means that this may hold too for complexity in subjects. But even this inclination is a hypothesis, and not a fact—as far as I am aware.

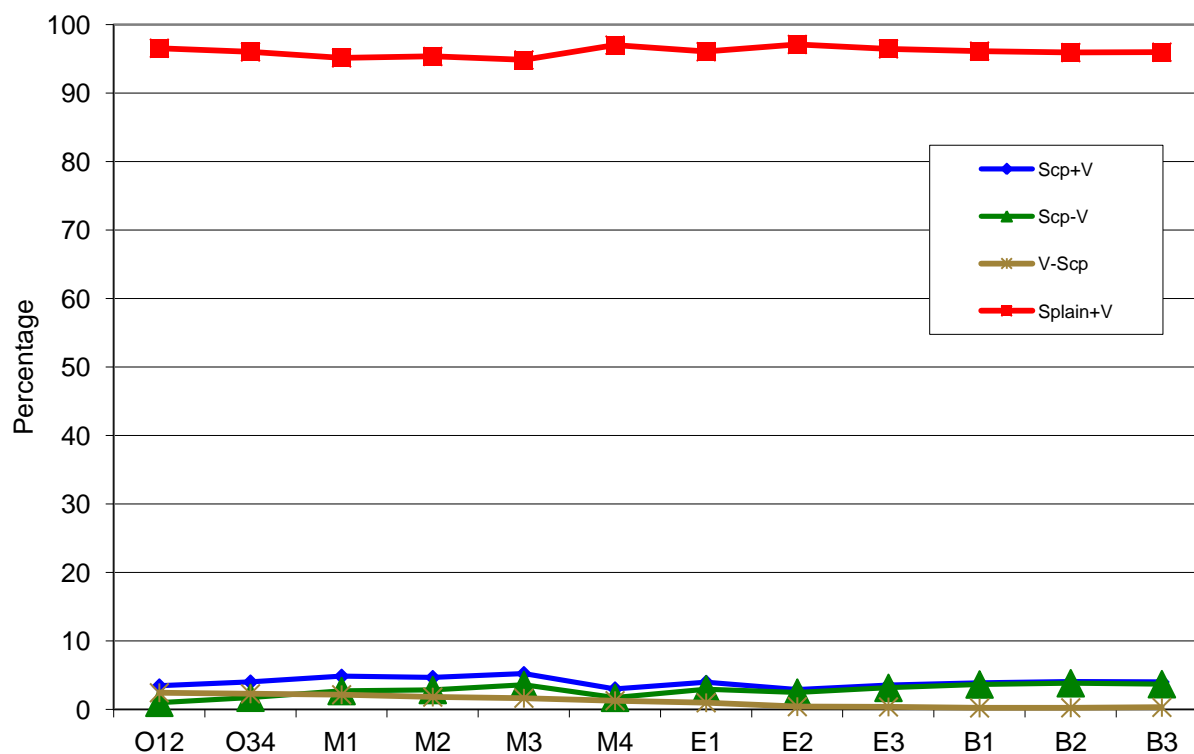
If we would have a look at relative-clause subjects, then it seems difficult to find an alternative to using a complex subject as in (2a). The alternative in (2b) feels rather awkward, and may even be misunderstood (the referent of *that man* may not necessarily be the same as *some man*). The alternative in (2c) is a bit of cheating: *my teacher* cannot really be regarded as subject. So this alternative is not valid. The fourth alternative in (2d) doesn't quite capture the same neutrality in meaning as (2a) does. It can be used if there is, for instance, if there is doubt about the identity of the person "walking over there".

- (2) a. _[NP-SBJ] The man [_{CP} who walks over there]] is my teacher.
b. _[NP-SBJ] Some man] walks over there. [_{NP-SBJ]} That man] is my teacher.
c. _[NP-SBJ] My teacher] is [_{NP-OB} the man [_{CP} who walks over there]].

- d. [NP-SBJ It] is [NP-OB my teacher] [CP-CLF who walks over there].

Results corpus research project “SubjectComplexity” provides insight into the change of the complexity of the subject through time. Figure (3) shows that the percentage of plain versus complex subjects does not change dramatically over time. Complex subjects, indicated by the “Scp+V” line, were rare in Old English (O1-4), but are almost equally rare in Modern British English (B1-3).

(3) Percentage of plain versus complex subjects diachronically.



3. Types of complex subjects

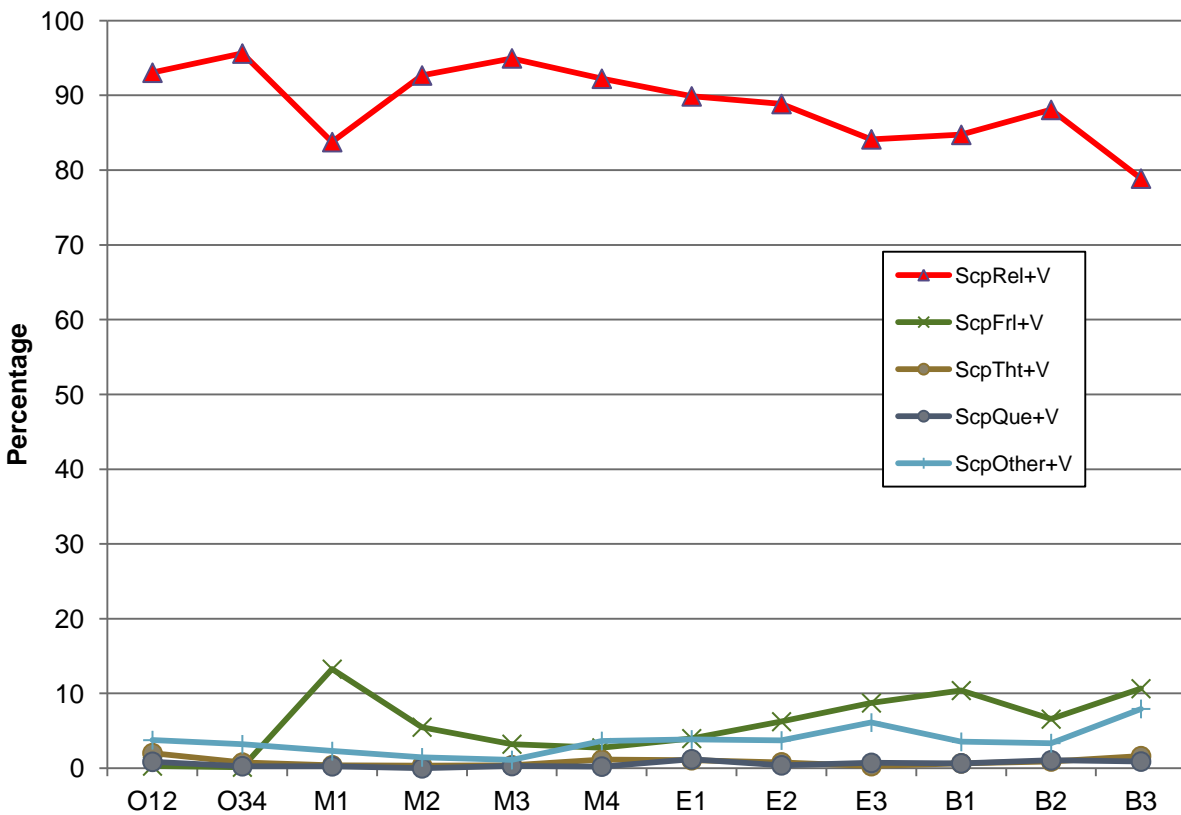
The previous chapter showed that complex subjects were and are a minority in English throughout time. But perhaps the *kind* of complex subjects have changed over time? The corpus research project “SubjectComplexity” considers 4 different CP types. The first type is a simple relative clause, as in (4a), which is indicated by CP-REL in the corpora. Free relative clauses can also serve as subjects, witness (4b). Rare throughout time indeed are subjects that contain a CP-THT, as illustrated by (4c). Likewise rare are subjects containing a question CP as in (4d). There is a fifth category of “other CP’s”, but this is a small category and it should not concern us too much.

- (4) a. [NP-SBJ The vast series of classical scholars [CP-REL that have written in the modern languages]] ought long before this time to have embodied whatever beauties can be passed on from the ancient literatures. [Behn-1878,377.277]
- b. [NP-SBJ [CP-FRL Whatever boys behave well in forwarding their companions]], deserve to be commended. [Barclay-1743,22.106]

- c. And [NP-SBJ the Ladie's saying [CP-THT I made him pay for his wife's wedding apparell]] is false. [Hoxinden-1650-E3-P2,168.32]
- d. And [NP-SBJ resouns of þe fend [CP-QUE wher Crist was boþe God and man]] marride hym, so þat he wyste neuere wer þis were sob or false; [CmWycSer,395.3050]

The relative frequency of these different types of complex subjects is illustrated in Figure (5). By far the majority throughout time consists of relative clause subjects. The M1 period can safely be disregarded or combined with M2, since there are only very few data from M1, which distorts the picture. What is perhaps interesting is the rise of the free relative subject starting from Middle English and stabilizing in Modern British English. Such subjects are tied to the *wh*-clefts.

(5) Types of complex subjects.



All in all the results so far do not seem to warrant *major* changes over time in the subject complexity.

4. The position of complex subjects

So far we have looked at complex subjects from different angles, but we have not seen any major changes. The percentage of complex versus plain subjects stays almost equal over time in the development of the English language. The previous chapter showed slight changes in the *types* of complex subjects. Relative clause subjects become relatively less frequent, and free relative subjects (related to the *wh*-clefts) become relatively more frequent.

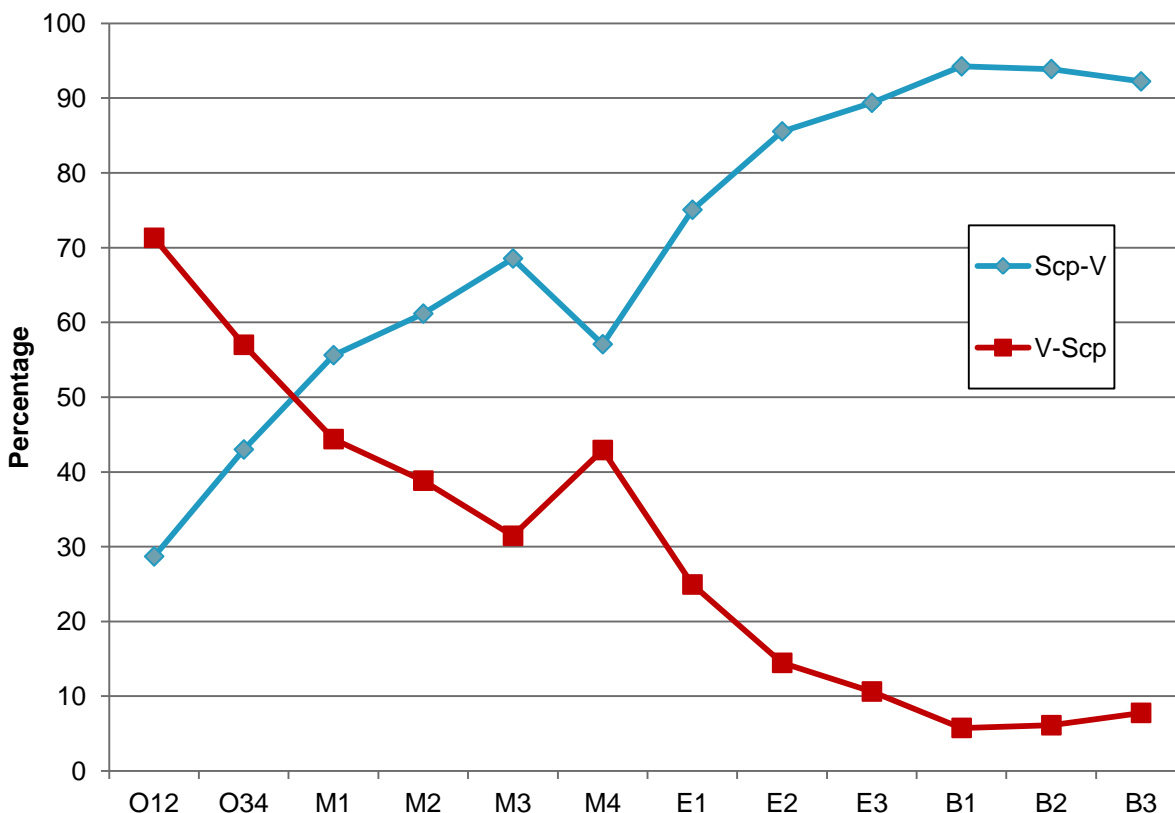
We will now turn our attention to the *position* of the complex subject, and see what changes are observed. I have chosen the position of the subject relative to the finite verb as a first rough yardstick for this research. This means that two different kinds of subject positions are possible.

The first one is the preverbal position, as illustrated by (6a). The second possibility is to have the subject after the finite verb, as in (6b)

- (6) a. and [_{NP-SBJ} he [_{CP-REL} that is conversant with the translations now accessible to the English reader]], [_V \$can] \$not be far from the kingdom of heaven. [Bain-1878,366.107]
- b. Various indeed [_V are] [_{NP-SBJ} the ways of working on these principles, [_{CP-REL} which one can no more enumerate, than describe the various tempers whence they proceed]]. [Barclay-1743,28.175]

The discussion referred to in the introduction might have led us to expect a *decrease* in preverbal complex subjects, but the measurements shown in Figure 7 tell us a completely different story. If we disregard the anomaly in M4, then there is a steady *increase* in preverbal complex subjects that works its way from Old English until early Modern English. The situation has come to a balance during the Modern British English period (1770-1910).

(7) The position of complex subjects relative to the finite verb.



The vast majority of complex subjects are relative clauses, followed in early Modern English by free relative subjects on a modest second place, as we have seen in section 3. That is why the change in complex subject *types* cannot be regarded as contributing to the steady change in complex subject *position*. The reason for this change in position must be sought in different areas.

5. Discussion

This paper has been concerned with complex subjects, that is, subjects containing a CP. Examples are relative clause subjects and free relatives. A fellow linguist commented that

complex subjects do not seem to fit in clause-initial position that well. This comment triggered me to investigate what happened diachronically to complex subjects in English. Corpus study showed that the percentage of complex versus simple subjects does *not* change significantly over time as the English language developed. There is, however, a slight change in the *types* of complex subjects occurring. By the end of the Middle English period the amount of free relative subjects increases slightly, and stabilizes during Modern English. By far the greatest change, however, occurs completely contra the expectations raised initially in this paper. There is a steady rise in complex subjects occurring *before* the finite verb over time. Old English starts with 30% complex subjects before the verb, and 70% following it. Modern British English shows more than a reversal. By that time 95% of the complex subjects occur before the finite verb, and a mere 5% follow it.

The reason for this change cannot be found in the change in complex subject *types*, which accounts for no more than 5%. The main reason for the change in complex subject position is perhaps related to the change in English syntax. The preverbal position used to be able to hold all kinds of constituents in Old English, a language structurally quite like modern German and Dutch. Present day English has a rather rigid SVO structure, and apparently the rigidity of that structure is such, that complex subjects are only rarely exempted by it.

6. References

TODO: add references.

7. Appendix

This appendix contains the details of the corpus research project “SubjectComplexity” used for the research above.

7.1 General information

Name: **SubjectComplexity_V1**

Author: Erwin R. Komen

Goal: Look at complex versus non-complex subjects

Comments: Complex subjects have a CP, whereas simple subjects don't.
We look first and foremost at:

- (1) main clauses
- (2) having a subject and
- (3) a finite verb

Look at the following subject types:

- (1) Simple. This subject has no CP
- (2) Complex. The subject has a CP
Finer distinctions are possible:
 - a. CP-THT
 - b. CP-REL
 - c. CP-FRL
 - d. Other CP (which are??)

Last change: 20-10-2010 6:30:07 (created: woensdag 20 oktober 2010 10:38)

Project type: Penn-psd

Queries: D:\Data Files\Corpora\CorpusStudio\Query

Output: D:\Data files\Corpora\CorpusStudio\Subject\Cs

Sources: D:\Data files\Corpora\English\psd\AllPeriods*.psd

Period Info: D:\Data Files\Corpora\CorpusStudio\Query\EnglishPeriods.xml
(changed:maandag 11 oktober 2010 9:06)

Parameters: Prec=2 Foll=1

7.2 Query construction

Line	Input	Query	Output	Result	Cmp	Exmp	Goal
1	Source	matSV	subSV	SV	-	-	Get main clauses containing a subject and verb
2	1/out	matScp+V	matScp+V	Scp+V	+	+	Main clauses with complex subject and finite verb
3	2/out	matScp-V	matScp-V	Scp-V	+	+	Main clauses with S[complex]...V
4	3/cmp	matV-Scp	matV-Scp	V-Scp	-	+	Main clauses with V...S[complex]
5	2/out	matScpRel+V	matScpRel+V	ScpRel+V	+	+	Main clauses with finite verb and subject containing a CP-REL
6	5/cmp	matScpFrl+V	matScpFrl+V	ScpFrl+V	+	+	Main clauses with finite verb and subject containing a CP-FRL
7	6/cmp	matScpTht+V	matScpTht+V	ScpTht+V	+	+	Main clauses with finite verb and subject containing a CP-THT
8	7/cmp	matScpQue+V	matScpQue+V	ScpQue+V	+	+	Main clauses with finite verb and subject containing a CP-QUE
9	8/cmp	matScp+V	matScpOther+V	ScpOther+V	-	+	Main clauses with finite verb and subject containing a CP that is not Rel, Frl or Tht
10	2/cmp	matSV	matSplain+V	Splain+V	-	+	Main clauses with simple subject and finite verb

7.3 Definitions

(Only definitions used in the queries are shown.)

matrixIP IP-MAT*

finiteverb BEI | BEP* | BED* | UTP | *HVI | *HVP* | *HVD* | *AXI | *AXP* | *AXD* | *MD | VBI | *VBP* | *VBD* | *DOI | *DOP* | *DOD* | NEG+BEI | NEG+BEP* | NEG+BED* | NEG+AXI | NEG+*AXP* | NEG+*AXD* | NEG+*MD | NEG+VBI | NEG+*VBP* | NEG+*VBD

subject \$subjectoe | NP-SBJ*

7.4 Queries

7.4.1 matSV

File: D:\Data files\Corpora\CorpusStudio\Query\matSV.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:

(1) a subject and

(2) a finite verb

These may be in any order

Changed: woensdag 20 oktober 2010 9:02 (created: maandag 11 oktober 2010 6:33)

Query:

```
node: IP-MAT*
add_to_ignore: \**
remove_nodes: f
print_indices: t
define: OE+MEU.def
```

```
query: (matrixIP iDoms subject) AND
       (matrixIP iDoms finiteverb)
```

7.4.2 matScp+V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScp+V.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:

(1) a subject and

(2) a finite verb

These may be in any order

Subject has a CP as descendant

Changed: woensdag 20 oktober 2010 9:25 (created: woensdag 20 oktober 2010 9:03)

Query:

```
node: IP-MAT*
add_to_ignore: \**
remove_nodes: f
print_indices: t
define: OE+MEU.def
```

```
query: (matrixIP iDoms subject) AND
       (subject Dominates CP*) AND
       (matrixIP iDoms finiteverb)
```

7.4.3 matScpRel+V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScpRel+V.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:
(1) a subject and
(2) a finite verb
These may be in any order
Subject has a relative clause CP

Changed: woensdag 20 oktober 2010 9:25 (created: woensdag 20 oktober 2010 9:09)

Query: node: IP-MAT*
add_to_ignore: **
remove_nodes: f
print_indices: t
define: OE+MEU.def

query: (matrixIP iDoms subject) AND
(subject Dominates CP*REL*) AND
(matrixIP iDoms finiteverb)

7.4.4 matScpFrl+V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScpFrl+V.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:
(1) a subject and
(2) a finite verb
These may be in any order
Subject has a free relative CP

Changed: woensdag 20 oktober 2010 9:25 (created: woensdag 20 oktober 2010 9:10)

Query: node: IP-MAT*
add_to_ignore: **
remove_nodes: f
print_indices: t
define: OE+MEU.def

query: (matrixIP iDoms subject) AND
(subject Dominates CP*FRL*) AND
(matrixIP iDoms finiteverb)

7.4.5 matScpTht+V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScpTht+V.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:

(1) a subject and

(2) a finite verb

These may be in any order

Subject has a that-CP

Changed: woensdag 20 oktober 2010 9:25 (created: woensdag 20 oktober 2010 9:10)

Query:

```
node: IP-MAT*
add_to_ignore: \**
remove_nodes: f
print_indices: t
define: OE+MEU.def

query: (matrixIP iDoms subject) AND
       (subject Dominates CP*THT*) AND
       (matrixIP iDoms finiteverb)
```

7.4.6 matScpQue+V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScpQue+V.q

Goal: Get all main clauses with a subject and a finite verb

Comment: Main clause contains:

(1) a subject and

(2) a finite verb

These may be in any order

Subject has a question CP

Changed: woensdag 20 oktober 2010 9:26 (created: woensdag 20 oktober 2010 9:16)

Query:

```
node: IP-MAT*
add_to_ignore: \**
remove_nodes: f
print_indices: t
define: OE+MEU.def

query: (matrixIP iDoms subject) AND
       (subject Dominates CP*QUE*) AND
       (matrixIP iDoms finiteverb)
```

7.4.7 matScp-V

File: D:\Data Files\Corpora\CorpusStudio\Query\matScp-V.q

Goal: Get all main clauses with a subject and a finite verb ordered S-V

Comment: Main clause contains:
(1) a subject and
(2) a finite verb
These are in the order of S...V
Subject has a CP as descendant

Changed: woensdag 20 oktober 2010 10:36 (created: woensdag 20 oktober 2010 10:36)

Query: node: IP-MAT*
add_to_ignore: **
remove_nodes: f
print_indices: t
define: OE+MEU.def

```
query: (matrixIP iDoms subject) AND
       (subject Dominates CP*) AND
       (matrixIP iDoms finiteverb) AND
       (subject Precedes finiteverb)
```

7.4.8 matV-Scp

File: D:\Data Files\Corpora\CorpusStudio\Query\matV-Scp.q

Goal: Get all main clauses with a subject and a finite verb ordered V-S

Comment: Main clause contains:
(1) a subject and
(2) a finite verb
These are in the order of V...S
Subject has a CP as descendant

Changed: woensdag 20 oktober 2010 10:37 (created: woensdag 20 oktober 2010 10:36)

Query: node: IP-MAT*
add_to_ignore: **
remove_nodes: f
print_indices: t
define: OE+MEU.def

```
query: (matrixIP iDoms subject) AND
       (subject Dominates CP*) AND
       (matrixIP iDoms finiteverb) AND
       (finiteverb Precedes subject)
```