Null Subject in English: Wonder if it exists?¹
Dawn Harvie
University of Ottawa

Although English is always claimed to be the prime example of a language in which every tensed clause must contain a subject, an examination of conversational English shows that sentences in which the subject is omitted regularly occur. Previous quantitative work on null subjects has been done for Spanish (Silva-Corvalan, 1982, 1994) and Portuguese (Paredes Silva, 1993), but apart from Core (1996) who largely focuses on discourse factors, there has been no quantitative work for English. This study examines the conditions on the occurrence of null subject in tensed main clauses in English discourse. The data were extracted from spontaneous tape-recorded conversations of Ottawa residents and were coded for several linguistic and discourse factors. Results show that switch reference and position in the clause contribute significant effects to the choice of null subject. Although many of the characteristics that have been put forward to account for the properties of pro-drop languages are not shared by English, these findings demonstrate that grammatical, social and processing constraints can act together to override seemingly categorical syntactic requirements for overtly filling subject position.

1.0 Introduction

Is modern English a pro-drop language? According to both prescriptive grammar and linguistic theory, no. Unlike true pro-drop languages, such as Spanish, modern English must have an overt subject, whether nominal or pronominal, in every finite clause. And yet, according to Mitchell & Robinson (1986), null subjects are in fact found in Old and Middle English, especially if the referent is the same as that of the preceding clause, as shown in (1).

(1) Yung of geres ase he was, Ø fleih awei into per wildernesse.  
"Young of years as he was, Ø fled away into the wilderness.  
(Visser, 1970)

According to prescriptive grammar, null subjects in modern English are possible, but are restricted to the deletion of identical subjects in conjoined sentences or chained clauses, as shown in example (2).

(2) I moved into this town 'n Ø got introduced to all these people.  
(103, 207B, 215)²

¹Thanks to Shana Poplack for permission to use data from the Sociolinguistics Laboratory Archives of Spoken Language Materials at the University of Ottawa. Thanks again to Shana Poplack, as well as James Walker, Michele Foley and everyone else involved in the Sociolinguistics Research Group for their help and useful comments on this paper.
²Numbers in parentheses represent speaker number, tape number and side, and counter number in the Sociolinguistics Laboratory Archives of Spoken Language Materials at the University of Ottawa.
While this constraint on the context of null subjects suggests that it should be highly restricted, in my own investigation of recorded spontaneous conversation in Ottawa English, I did find examples of null subjects, one of which is shown in (3).

(3) Ø watch a movie or something, you know
    (246, 212 A, 055)

That such examples exist is not surprising from a variationist point of view, since we should expect the natural discourse of speakers to vary from what the prescriptive grammars say. But it is unexpected, given that English does not fulfill the criteria proposed for null subject languages by theoretical linguists.

Theoretical analyses of null subject restrict themselves to pronominal subjects and distinguish between referential null subjects with an identifiable antecedent on the one hand, and non-referential or expletive subjects on the other. In this literature, it is claimed that referential null subjects are allowed because the language has rich agreement morphology, as argued by scholars such as Chomsky (1982), and Rizzi, (1982, 1986). This requirement means that the verbal inflectional paradigm must uniquely identify each person, which it does not for English. According to Visser (1970), this is also true of Old English. This analysis does account for the presence of null subjects in languages like Spanish and Italian, but it is problematic for languages such as Japanese and Chinese (Huang, 1984), which show no inflection for number or person on the verb at all. In order to account for this discrepancy, Jaeggli and Safir (1989) propose the principle of morphological uniformity, as shown in (4), which requires the conjugational paradigm of null subject languages to consist uniformly of either bare or affixed stems.

(4) The Principle of Morphological Uniformity
    An inflectional paradigm P in a language L is morphologically uniform iff P has either only underived inflectional forms or only derived inflectional forms.
    (Jaeggli & Safir, 1989: 30)

They cite English as an example of a language with a mixture of morphologically complex and bare stems, since it standardly marks agreement only on the third person singular. Therefore, English does not qualify as a null subject language in either of the above frameworks.

There are some "semi-pro-drop" languages, such as Icelandic and Faroese, which allow a null variant only for non-referential or expletive subjects. Since,
according to theoretical analysis, English subjects must be overt, the grammar requires that the subject position be filled, and hence expletive or non-referential subjects would be inserted. Motivated solely by the assumption that English is not a pro-drop language, we would expect such semantically void elements to be potential null subjects.

Following the theoretical assumption that subject expression is variable only in “officially” pro-drop languages, variationist studies of this phenomenon have generally concentrated on such languages. For example, quantitative studies of null subject phenomena for spoken Spanish are found in Silva-Corvalán (1982), and Cameron (1993, 1996), for written Brazilian Portuguese in Paredes Silva (1993) and for written Arabic in Parkinson (1987). However, Cote (1996), and Lattey (1980), which study null subject in English, are notable exceptions to this trend.

Variationist studies of null subject have considered both linguistic and discourse factors. Given the importance of subject-verb agreement in theoretical analyses, verb form and the person and number features of the subject are important factors in most of these studies. Additional linguistic factors include the referential or non-referential nature of the subject, clause type, and pre- and postverbal position. Common discourse factors studied include switch reference, contrast, and turn position.

Although, as I mentioned above, English is considered a non-pro-drop language, it is generally recognized that imperatives and other written styles, such as that used in diaries, postcards and informal letters, generally allow null subjects, as Haegeman (1990) has shown. However, as was shown in (3), referential null subject does exist outside these contexts in spoken English.

2.0 Data and Method

The data used in this study were extracted from the Sociolinguistics Laboratory Archives of Spoken Language Materials at the University of Ottawa. These archives consist of tape-recorded, spontaneous conversations with speakers of English in the Ottawa region collected by students taking Urban Dialectology I (LIN 3142) at the University of Ottawa. For the purposes of this study, I selected 14 working-class speakers with a low rating in the linguistic marketplace (Sankoff & Laberge, 1978). There were 8 male and 5 female speakers, ranging in age from 15 to 45.

I extracted a total of 697 tokens. Since null subject is such a rare phenomenon, I had to use a method of extraction which would result in a
representation of null subject tokens sufficient to permit a multivariate analysis by GOLDVARB (Rand and Sankoff 1990). I first identified clauses in which a null subject occurred. Wherever possible, I then selected contextually-related adjacent clauses which may or may not have had null subjects. The result was a total of 224 null subject tokens and 471 overt subject tokens.

Every token was coded for four linguistic factor groups: subject type, clause type, negation, and position in clause; and three discourse factors: switch reference, contrast and turn position. Given the importance of verbal agreement, the subject was coded for person. Items, such as "you know" and "I guess", were excluded because they have been lexicalized. In coding for subject type, I initially coded for person, together with number (for example, first person singular, first person plural) but due to insufficient distribution of all subject types, I later coded for first person, second person, third person and expletive. Tokens of generic second person, as in (5a), were included with non-generic second person, as in (5b), because there were so few non-generic tokens (N=6).

(5a) You take your girlfriend on a Sunday afternoon
(239, 417A, 204)

(5b) You should take her out to the different bars they have here in Ottawa
(246, 212A, 303)

Third person subjects were coded as expletive or referential, in an attempt to capture the important distinction I mentioned above.

Clauses were coded as either conjoined or non-conjoined. This factor group was coded to determine whether there is any difference in the factors conditioning null subject between these two types of clauses, since it is quite acceptable to have a null subject in a conjoined clause.

Exclusions include interrogatives, such as (6), since I found no variation in this context; subjects are categorically overt.

(6) What are you doing in this house?
(Interviewer for 239, 417A, 001)

Due to the possibility of double deletion, I excluded all contexts of subject-aux deletion, including yes/no questions. I also excluded subordinate clauses, in which, as shown in (7), both the subject and subordinate conjunction can be deleted.
(7) Yeah make sure ... tell Pawpa Ø [that] Ø [I] hope he feels better.  
(Hopper, 1992)

Relative clauses were excluded on the basis that the relative pronoun, according to theoretical analyses, undergoes movement from the subject position of the clause. Following Cote (1996), I excluded instances of repetitions, where a speaker repeats another speaker's words either exactly or with some small change, (8), and self-corrections, (9), where a speaker corrects his/her sentence without repeating the subject. While imperatives are not entirely without variation, see (10), they were excluded because the unmarked form is null subject.

(8) A: It was just the first of September  
(239, 417B, 066)  
B: Ø [it was] just the first of September  
(Interviewer for 239, 417B, 067)

(9) Ah, I wouldn't want to be anywhere else wouldn't want to be ah (239, 418A, 101)

(10) Ø stop that!  
You stop that!  
(Cote, 1996)

Although negation was not included as a factor in the other studies of null subject that I consulted, I coded for the presence or absence of negation based on the relatively frequent occurrence of sentences such as (11).

(11) Ø haven't thought of stuff like that.  
(239, 418B, 180)

Under the assumption that the sentence-initial position would favor subject deletion, I also coded for the position of the subject in the sentence. This factor group investigates whether or not hesitations, subordinate clauses, or other elements that precede the subject might condition null subject.

The discourse factors of switch reference, contrast and turn position have been significant in other variationist studies of pro-drop languages. As in (12), a token was coded same if the referent of the subject under consideration was the same as the subject referent of the previous clause and switch if the referent of the subject of the previous clause was different.
(12) It's just, you know, it's pretty boring. You (switch) got there. You (same) eat. You (same) dance. Ø (same) drink. Stuff like that.
(246, 212A, 008)

I followed Silva-Corvalan’s (1982) method of coding for contrast. She codes a subject NP as “focus of contrast” if the referent of the NP in question stands in opposition to a closed number of clearly identifiable alternatives. The difference between alternatives in opposition and a list of alternatives is that, in the former, only one of the alternatives in a contrasting situation may be chosen as the right one. This difference is crucial to the definition of contrast (Silva-Corvalan 1982). Although this factor exhibits some overlap with switch reference, I coded it separately, but was later forced to discard it, since subjects in contrast were always overt. Recoverability may also be a consideration: in pro-drop languages, the subject can often be recovered from the verbal morphology, whereas this is not generally possible in English.

Finally, following Cote (1996), the discourse position of the clause was coded for the beginning, middle, and end of a turn, as well as for single-utterance turns. Based on initial results, this was later re-coded as turn-initial versus non-turn-initial.

Table 1 shows the results of multivariate analysis. Firstly, note that switch reference was selected as most significant, with same-referent subjects contributing the greatest effect to the occurrence of null subject, with a weight of .57. This finding coincides with that of Cameron (1993) for Puerto Rican and Madrid Spanish, Silva-Corvalan (1982) for Mexican-American Spanish, and Parkinson (1987) for Egyptian Arabic. Thus, this factor group is significant for both pro-drop and non-pro-drop languages.
Table 1: Contribution of factors selected as significant to the probability of null subject in Ottawa English with percentages and raw numbers

<table>
<thead>
<tr>
<th>Factor Group</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same as subject of previous clause</td>
<td>.57</td>
<td>38</td>
</tr>
<tr>
<td>Switched from subject of previous clause</td>
<td>.36</td>
<td>20</td>
</tr>
<tr>
<td>Range</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Position in Clause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>.54</td>
<td>36</td>
</tr>
<tr>
<td>Non-initial</td>
<td>.40</td>
<td>23</td>
</tr>
<tr>
<td>Range</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Corrected mean: .308</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total: 697</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor groups not selected: Subject Type, Clausal Type, Negation and Turn Position

Switch reference can be related to Grice’s discourse maxims of quantity, as shown in (13)

(13)  *Grice’s Maxims of Quantity*

a) Make your contribution as informative as is required (for the current purposes of the exchange).

b) Do not make your contribution more informative than is required

Therefore, if a subject has just been specified, it is not necessary to repeat it. This discourse factor group, therefore, seems to account for, at least in part, the dropping of the subject in English. In fact, considering its importance to both pro-drop and non-pro-drop languages, it may possibly be a universal discourse principle for null subject. Note that switch reference was also considered relevant to the occurrence of null subject in Old English, in which a subject could be dropped if it was unambiguous.

The other factor group selected as significant was the position in the clause. Sentence-initial subjects contribute the greatest effect with a probability of .54. It seems reasonable to expect that subjects that are not preceded by other material would be easier to drop. This may have more to do with phonological or prosodic constraints that I have not investigated, since current phonological theory has paid much attention to edge effects (McCarthy, 1993; McCarthy & Prince, 1993), especially those involving prosodically weak elements, such as pronouns.

Although not chosen as significant, the subject type is worth discussing, because of the findings of other studies. The percentages for this group, shown in Table 2, indicate that third and second person are dropped approximately 35% of the time, with expletives and first person following at approximately 30%.
Table 2: Distribution of null subject according to subject type

<table>
<thead>
<tr>
<th>Subject Type</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third person</td>
<td>35</td>
<td>211</td>
</tr>
<tr>
<td>Second person</td>
<td>34</td>
<td>58</td>
</tr>
<tr>
<td>Expletive</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>First person</td>
<td>29</td>
<td>401</td>
</tr>
<tr>
<td>Total</td>
<td>697</td>
<td></td>
</tr>
</tbody>
</table>

Cote (1996) found that expletives were dropped most often. This would follow from the assumption that expletives are semantically void syntactic place-holders. Also recall that semi-pro-drop languages allow the expletive only to be dropped. Perhaps expletives were not found to be dropped more than pronouns in my study because there were only 20 tokens with expletives. Interestingly, and paralleling my results, Parkinson (1987) found that for Egyptian Arabic, third person was most frequently null, followed by second person. However, he provides no explanation for this pattern. For my English results, the marking of third person singular with -s may be a contributing factor to the high rate of deletion in third person. This hypothesis would be worth investigating in future studies. We would also expect that generic you would have a relatively high rate of deletion, because of the essentially non-specific nature of these subject types.

As part of this study, I also examined non-conjoined and conjoined sentences for any differences in the conditioning of null subject between these two types of sentence. This factor group was not found to be significant. But the percentages for conjoined versus non-conjoined sentences, shown in Table 3, demonstrate that, with the exception of subject type and position in clause the same ranking of constraints applies.
Table 3: Distribution of null subject across main and conjoined sentences

<table>
<thead>
<tr>
<th>Subject Type</th>
<th>Main %</th>
<th>Main N</th>
<th>Conjoined %</th>
<th>Conjoined N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expletive</td>
<td>38</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Third person</td>
<td>38</td>
<td>150</td>
<td>27</td>
<td>60</td>
</tr>
<tr>
<td>First person</td>
<td>28</td>
<td>334</td>
<td>36</td>
<td>67</td>
</tr>
<tr>
<td>Second person</td>
<td>26</td>
<td>47</td>
<td>73</td>
<td>11</td>
</tr>
<tr>
<td>Switch Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>35</td>
<td>353</td>
<td>47</td>
<td>98</td>
</tr>
<tr>
<td>Switch</td>
<td>23</td>
<td>194</td>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>Negation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>38</td>
<td>112</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>Absent</td>
<td>30</td>
<td>441</td>
<td>34</td>
<td>127</td>
</tr>
<tr>
<td>Position in Clause</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>36</td>
<td>385</td>
<td>34</td>
<td>118</td>
</tr>
<tr>
<td>Non-initial</td>
<td>21</td>
<td>168</td>
<td>36</td>
<td>25</td>
</tr>
<tr>
<td>Turn Position</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-initial</td>
<td>32</td>
<td>454</td>
<td>34</td>
<td>139</td>
</tr>
<tr>
<td>Initial</td>
<td>28</td>
<td>99</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Totals</td>
<td>553</td>
<td>143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This leads me to believe that they behave in a similar way with respect to null subject, the main difference being that, syntactically, subject deletion is allowed in conjoined sentences. Interestingly, there were only four expletive tokens in conjoined sentences and they were all overt. There were also only nine subjects that were preceded by hesitations, or something like a subordinate clause. I did not consider 'and' to be the same as a hesitation or subordinate clause, otherwise all subjects in conjoined sentences would be considered to be non-initial in the sentence. This decision needs further consideration. Future work could investigate this and whether discourse factors apply to conjoined sentences to the same degree that they apply to non-conjoined sentences.

3.0 Conclusion

Before concluding, I would like to mention two difficulties that I encountered in the course of this study.

Firstly, people in the speech community differ greatly in their use of null subject in both conjoined and non-conjoined sentences. There are a number of possible reasons for this, none of which has been investigated. For example, it may be related to social factors, such as stigmatization, though again this is an area for further investigation.

Finally, finding what the prescriptive and theoretical grammarians define as a “sentence” in naturally-occurring, spontaneous discourse is a challenging task.
Conversation is full of stops and starts, hesitations, interruptions and other such phenomena, which means that defining the boundaries of sentences for the purpose of extracting tokens leads to a large number of potential tokens which must be discarded because of their inherent ambiguity.

Null subject does indeed exist in English, and it has for centuries. Visser (1970) makes it clear that, in Old and Middle English, the use of the subject pronoun was the rule rather than the exception. Mitchell and Robinson (1986) state that subjects in Old English were sometimes dropped if the referent was the same as that of the previous clause. Therefore, the constraints that held for null subject in Old and Middle English appear to be the same constraints which hold for null subject in present-day English.

More interestingly, the constraints that hold for English are also important for pro-drop languages. Exactly what this means is open for debate. I do not believe that it means that English is a pro-drop language, or even a semi-pro-drop language, but it does hint at the universal nature of this phenomenon. The importance is not the difference between pro-drop and non-pro-drop languages, but rather the similarity of the factors conditioning null subject in all languages in which it occurs.

Finally, the disparity between daily experience and textbook grammar exposes the very real gap between natural language and the prescriptive rules for language. This study is one of a very few investigating this gap for null subjects in English and, as such, requires elaboration. Further study of null subject, and especially constraints of discourse on null subject, would prove beneficial, not only to variationists, but also to theorists, discourse analysts, and - who knows? - even to prescriptive grammarians.

References


