English Language and Linguistics 3.1: 117–45. © Cambridge University Press 1999.

Subject ellipsis in embedded clauses in English[1]

LILIANE HAEGEMAN AND TABEA IHSANE
University of Geneva
(Received 8 April 1998; revised 19 November 1998)

This paper examines subject omission in English finite clauses. Contrary to what is claimed in Haegeman (1997), it is shown that embedded subject omission is attested in diary-style registers. These data pose a problem for the theory of empty categories in the *Principles and Parameters* framework. The main body of the paper describes the contexts in which subjects of embedded finite clauses are omitted. No syntactic constraints have been identified. Based on the observation that the omission of embedded subjects coincides with pronoun omission of reflexives, we tentatively suggest that in specific registers pronoun ellipsis is licensed by a specifier–head relation with a head carrying agreement features.

1 Scope of the paper

Haegeman (1997, forthcoming) discusses the omission of the subject in English finite sentences. The core data discussed in these papers are those in (1).

(1) (a) *Have bought a newspaper.
(b) 'Couldn't get herself a man, so she was always showing what she could get. Wonder who'll get her money? Won't be me, thought, not so likely.' (Ruth Rendell, *A sleeping life*. 1994: 87–8) (from Haegeman, 1997: 236)
(c) Have done 110 pages (Virginia Woolf, *Diary*. 1985: 33) (from Haegeman, 1997: 236)

English does not generally allow nonovert subjects in finite clauses: (1a) is usually considered to be ungrammatical.[2] However, certain registers allow finite sentences with nonovert subjects. These registers concern 1. informal spoken English as in (1b) (see Schmerling, 1973; Thrasher, 1977; Lawrence, 1996) and 2. written registers in which pressure of economy brings about certain types of ellipsis as in (1c) (Haegeman, 1990a, 1990b, 1997, forthcoming). The relevant written register is exemplified by the diary writing in (1c); other types of abbreviated English (e-mail messages, notes (cf. Janda, 1985), postcards, information on commercial packaging,

[1] An earlier paper on the same topic, 'Embedded null subjects in English diaries', was cowritten with Tatiana Koutsandréou.

Warmest thanks to Helen Fielding for discussing the data with us and to Helen Lawrence for generously making available her material. Thanks also to Bas Aarts, David Adger, Enoch Aboh, Margaret Clayton, Siobhan Cottell, Cecile de Cat, Wlad Godzich, Eric Haeberli, Stephen Harlow, Liz McCoy, Bernadette Plunkett, Anthony Warner, Richard Waswo and two anonymous reviewers of *English Language and Linguistics* for useful comments. Thanks to Johan Haegeman and Sylvie Ferioli for help with the French material. The usual disclaimers apply.

[2] This statement is a simplification. See section 2.1.3.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

telegraphic styles (Barton, 1998, encyclopedia entries etc.) also show this property (see Haegeman, 1997; Horsey, 1998).

Haegeman's analysis (1997, forthcoming) is based on the observed similarities between the distribution of null subjects in spoken (1b) and written English (1c), and that of null subjects in the early English production data analysed, among others, in Hyams (1986), Valian (1991), and Rizzi (1994, 1995). Adopting Rizzi's (1994, 1995) analysis of the early null subject, Haegeman proposes that the adult null subject in (1b) and (1c) is an empty category in the highest NP position of the root clause. The empty category lacks a clause-internal antecedent and is identified in the discourse.

Based on Rizzi (1994, 1995), she proposes that an antecedentless empty category can only survive in a context in which no antecedent position is available, i.e. precisely in the highest NP position of the clause. This leads to the prediction that the adult null subject, like the early null subject, is excluded from embedded clauses. (1d)–(1g) are predicted to be ungrammatical. For the spoken register both Schmerling (1973: 583) and Thrasher (1977: 29) assert that subject deletion is a root phenomenon.

(1) (d) *I think will leave.
(e) *I will see if can get into the office.
(f) *He smiled at me when walked past.
(g) *This is what gave him.

For further details see Haegeman (1997, forthcoming). Horsey (1998) offers a minimalist reinterpretation of the analysis.

The present paper re-evaluates the analysis in Haegeman (1997, forthcoming) in the light of the observed distribution of the null subject in diary extracts. While part of our findings support Haegeman's analysis, we also present data which are problematic for that analysis. Specifically, embedded null subjects (in (1d–g)), from now on ENS, though rarer than root null subjects (in (1b–c)), from now on RNS, are found in diary extracts.

The paper is organized as follows. Section 2 summarizes the earlier analysis which predicts that null subjects in English finite clauses are a root phenomenon. Section 3 shows that the prediction is incorrect and that ENS are attested in the relevant written register. Section 4 describes in detail the distribution of the ENS. Section 5 shows that the ENS is problematic for the theory of empty categories as developed in the generative framework. This section explores an approach in terms of pronoun deletion, and the problems that this raises. Section 6 briefly discusses spoken English and section 7 summarizes the paper.

2 Nonovert subjects in English diaries

2.1 Left-peripheral null subjects and the theory of empty categories

The English empty subjects in (1b–c) do not fit easily into the classification of empty categories in the generative literature (cf. Chomsky, 1981, 1982), where four types of

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

nonovert categories are distinguished as illustrated in (2) (see Haegeman, 1997: section 2.4.1). We restrict our attention to subject-empty categories, for reasons of space.

(2) (i) Trace of A-movement (A-trace)[3]
Nelson$_i$ seems [$_{IP}$ t$_i$ to be hungry].
(ii) Trace of A′-movement (A′-trace)
(a) Who$_i$ do you think [$_{CP}$ t$_i$ [$_{IP}$ t$_i$ will get here first]]?
(b) Nelson$_i$, I don't think [$_{CP}$ t$_i$ [$_{IP}$ t$_i$ will get here first]].
(iii) *pro:* the nonovert pronoun in finite clauses of so-called *pro*-drop languages
pro parlo italiano (Italian).
speak-1sg Italian
'I speak Italian.'
(iv) PRO, the understood subject of nonfinite embedded clauses
[PRO to feed Nelson regularly] is important.

Because the content of nonovert categories must be recoverable, empty categories are subject to an identification condition. For English, the identification of empty categories takes place via an antecedent in the clause. Thus, for instance, the A-trace in the subject position of the embedded nonfinite clause in (2i) is identified by its antecedent, *Nelson*, in the matrix subject position. We will briefly compare the RNS in (1b) and (1c) with the empty categories in (2).

2.1.1 A-traces

A-traces are empty categories whose antecedents occupy A(rgument) positions. Subject A-traces are restricted to nonfinite clauses (3a). A-traces do not alternate with overt subjects (3b). They take as their antecedent the subject of an immediately dominating clause.

(3) (a) *Nelson$_i$ seems t$_i$ is hungry.
(b) *Nelson$_i$ seems *he$_i$* to be hungry.

The RNS in (1b–c) differs from the A-trace in that it occurs in a finite clause, where it alternates with an overt subject.

2.1.2 PRO

The empty category *PRO* (2iv) represents the understood subject of nonfinite embedded clauses:[4]

(4) (a) The cat expects [PRO to get regular meals].
(b) [PRO to be fed regularly] is important for the cat.
(c) When [PRO hungry], the cat comes into the kitchen.

Like the A-trace, discussed above, PRO is excluded from finite clauses (5a–c) and it does not alternate with overt subjects (5d).

[3] A-movement targets a position which may be associated with a grammatical function, essentially a subject position. A′-movement targets a position which is not associated with a particular grammatical function, such as the positions in the periphery of the clause to the left of the subject.

[4] Unlike the A-trace, PRO does not require an antecedent, as shown by (i):

(i) PRO to talk to her is a mistake.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(5) (a) *The cat expects [$_{CP}$ [$_{IP}$ PRO will get regular meals]].
(b) *[$_{CP}$ That [PRO should be fed regularly]] is important for the cat.
(c) *When [PRO is hungry], the cat comes into the kitchen.
(d) *[Us/we to be fed regularly] is important for the cat.

As it occurs in finite clauses, and as it alternates with an overt subject, the null subject in (1b–c) cannot be analysed as *PRO*.

2.1.3 A′- traces

A-traces and PRO occur in nonfinite clauses. They do not alternate with an overt subject and they may take a higher subject as their antecedent. An A′-trace is a trace of movement to an A′-position, i.e. a non-argumental position, typically in the left periphery of the clause. In (6a) the antecedent, i.e. the moved constituent, is a *wh*-phrase, in (6b) it is a topic.

(6) (a) Who$_i$ does Mary think [t$_i$ will get here first]
(b) Nelson$_i$, he doesn't think [t$_i$ will get here first]

As shown by these examples, the A′-trace may occupy the subject position of a finite clause.

The A′-trace cannot take a higher subject as its antecedent. In (6a), the matrix subject *Mary* cannot be interpreted as the antecedent of the nonovert embedded subject; *Mary* and *who* are not coreferential. In (6b), the embedded A′-trace takes as its antecedent the fronted topic *Nelson* and not the matrix subject *he*. Again, *Nelson* and *he* cannot be coreferential.

An A′-trace cannot remain antecedentless:

(6) (c) *He doesn't think [t$_i$ will get here first]

The antecedent of an A′-trace may, under restricted conditions, be nonovert. In (7a) the relative pronoun is overt (*who*); in (7b) a nonovert analogue of *who* is postulated (cf. Lasnik and Stowell, 1991). The nonovert pronoun is represented as OP (for nonovert operator). The nonovert operator has to be identified. In (7b) this is achieved by the adjacent head of the relative clause, *a friend*.

(7) (a) I need a friend [who$_i$ [I can trust t$_i$]].
(b) I need a friend [OP$_i$ [I can trust t$_i$]].

For other languages, a null operator analysis is also advocated to account for empty categories which are identified by a discourse antecedent. In the Chinese example (8a), the direct object of *kanjian* ('see') is understood to be an entity which is salient in the discourse. (8a) does not contain a clause-internal antecedent for the nonovert object. It is postulated that the object of *kanjian* is a nonovert topic which undergoes leftward movement and leaves an A′-trace in its base position. This is represented in (8b) (Huang, 1984, 1989). For similar data in Portuguese see Huang (1984), Raposo (1986).

(8) (a) Zhangsan xiwang [Lisi keyi kanjian]
Zhangsan hope Lisi can see
'Zhangsan hopes that Lisi can see her/him/them.'

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(b) [OP_i Zhangsan xiwang [Lisi keyi kanjian t_i]].

Haegeman (1990a, 1990b) and Bromberg and Wexler (1995) took the RNS in the special spoken and written registers (1b,c) to be an A′-trace of nonovert topic-antecedent in the left periphery of the clause. However, one argument against identifying the null subject as trace of a nonovert topicalized constituent concerns its interpretation. Though the large majority of nonovert subjects in the special registers are interpreted as referential, in (9) the nonovert subjects are nonreferential. Topicalization is incompatible with nonreferential NPs, as shown by the ungrammaticality of French (10).

(9) (a) Rained in the night, wind, rain and hail (Smart, *Necessary secrets*. 1991: 27)
(b) Is raining. ('The diary of Emma D. May', *Independent on Sunday*. 21 June 1998: 11)

(10) *Lui, [il pleut]
he, he rains

2.1.4 The nonovert pronoun

In the so-called *pro*-drop languages like Italian (cf. Rizzi, 1986, 1993; Haegeman, 1997), a nonovert pronominal, represented as *pro*, is postulated as the nonovert subject in apparent subjectless finite clauses such as (11).

(11) (a) [*pro*] lavorano nel giardino
[*pro*] work-3pl in the garden
'They are working in the garden.'
(b) I ragazzi cantano quando [*pro*] lavorano
The boys sing when [*pro*] work-3pl
'The boys sing when they work.'

The content of null-subject pronouns must be identified. In Italian this is achieved by the rich verbal inflection. The Italian verbal paradigm is 'rich' in that it has one form per person/number combination, allowing for the identification of the null pronoun as its subject. *Lavorano*, in (11), for instance, is unambiguously identified as the third person plural. The paradigm of the English verb *work* would not allow for the recovery of the subject.

The RNS in diaries tends to be interpreted as first- or third-person pronouns, singular and plural. (12) provides some examples. The lack of second-person interpretations is presumably due to the fact that diary writing rarely addresses an interlocutor directly.

(12)

	person/number	example
(a)	1sg	Don't share sorrow with Ted (Plath, *Journals*: 289)
(b)	3sg	Here, studies under [David] Daiches (Plath, *Journals*: 126)
(c)	1pl	Climbed Rue Vieuville (Plath, *Journals*: 136)
(d)	3pl	Sat there for 3/4 hour (Woolf, *Diary*: 334)

On the basis of these interpretations, one might propose that the diary RNS in (1b–c) is an instantiation of the nonovert pronoun found in *pro*-drop languages. A *pro*-drop analysis of the null subject in the diary-style writing raises both conceptual and empirical problems.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

If the null subject in the special registers (1b,c) were evidence for a resetting of the *pro*-drop parameter, one would have to conclude that English speakers/writers have two (mostly overlapping) grammars: one in which the *pro*-drop parameter is switched off, the 'core grammar', and a second one in which the parameter is switched on, the 'peripheral grammar'. The peripheral grammar can arguably be acquired on the basis of the overt evidence of subject omission in spoken English (see Thrasher, 1977; Schmerling, 1973; Haegeman, 1997).

Given a parameter resetting analysis to account for this kind of language-internal variation, one would expect to find other kinds of language-internal variation due to parameter resetting. Haegeman (1987) proposes that this is the case, and that (13), in which the relative pronoun *which* has been extracted from the adjunct clause introduced by *before,* is the result of the resetting of the subjacency parameter, the parameter which determines the distance of extraction (see Haegeman, 1987 for details).[5]

(13) plays which$_i$ [the committee itself had to turn the public out of the room and close the doors [before it could discuss t_i]] (G. B. Shaw, *The Doctor's Dilemma*. 1911: 309; Haegeman, 1987: 215)

However, not just any random type of resetting seems to be available in register-related variation. We do not seem to find languages with verb–object order in one register, and which display object–verb order in another, and it is unlikely that we will find such a pattern. The question of how to delimit language-internal parametric variation is not addressed in Haegeman (1987). For a theory of register-based parameter resetting to work, one would have to be able to differentiate in a principled way between those parameters which will allow register-based resetting and those which don't, and one would have to specify for each case how resetting is triggered by the evidence.

An empirical problem for the *pro*-drop analysis of the RNS in (1b–c) concerns the identification of the null subject. If certain spoken and written registers allow an Italian-style nonovert subject pronoun, the question arises as to how this pronoun is identified, the verb inflection in these registers being no richer than that of other registers. For (14a), the verb form *am* does unambiguously identify the subject, but not all diary null subjects co-occur with such verb forms. *Have* in (14b) and *write* in (14c) are compatible with first-person singular subjects as well as with second-person subjects, with first-person plural subjects and with third-person plural subjects. The modals in (14d,e) and the past tense verbs in (14f,g) are compatible with subjects of any person/number combinations.

(14) (a) Am reading the book of Job. (Plath, *Journals*: 290)
(b) Have been awake since black 3.30. (Plath, *Journals*: 154)
(c) Write down the passing thought. (Plath, *Journals*: 165)
(d) Couldn't even read. (Plath, *Journals*: 108)

[5] Haegeman (1987) fails to discuss what kind of overt evidence would be required to trigger this parameter resetting.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(e) May drive to Paris with him. (Plath, *Journals*: 127)
(f) Got up as brash, nerve-raking alarm ground off at 6. (Plath, *Journals*: 54)
(g) Cried yesterday morning. (Plath, *Journals*: 288)

Subject omission in Italian type *pro*-drop languages is a pervasive phenomenon and, crucially, takes place in embedded clauses as shown by (11b). In such languages, subject omission also takes place in *wh*-questions and in *yes*/*no* questions:

(15) (a) (Quando) tornerà [*pro*]?
(when) will-come-back [*pro*]
'(When) will he/she return?'

Again, Haegeman (1997) argues that English subject omission is not possible in embedded clauses (1d–g), and in *wh*-questions or *yes-no* questions (15b).

(15) (b) *(When) will [ec] invite me again (Haegeman, 1997: 239)

A *pro*-analysis of subject omission in (1b,c) would have to account for the distributional constraints on the subject ellipsis.

For completeness' sake note that the nonovert pronoun is not only postulated for languages with rich inflection. It has been proposed that languages without any overt agreement at all also may exhibit *pro*. Huang (1989) argues that the null subjects in the Chinese finite clauses in (16) are instantiations of *pro*, in spite of the absence of any overt verb inflection. The null subject alternates with an overt pronoun.[6]

(16) (a) *pro*/Ta kanjian ta le (Huang, 1989: 187)
pro/he see him ASP
'He saw him.'
(b) Zhangsan shuo [*pro*/ta lai le] (Huang, 1989: 188, his (5))
Zhangsan say *pro*/ he come ASP
'Zhangsan said that he came.'

Transposing a Chinese type *pro* to the English sentences (1b,c) raises at least two problems. First, even though impoverished, English has agreement morphology, and the third-person singular ending remains available in informal spoken English and in diary-type registers. Furthermore, Haegeman (1997) signals that the diary null subject occurs in French, a language which clearly exhibits inflectional endings on the verb.

Secondly, as shown by (16b) the Chinese null subject occurs in embedded clauses, a pattern assumed by Haegeman (1997) to be ungrammatical both for the spoken register (1b) and for the written register (1c).

[6] The absence of overt agreement is not sufficient, though, to allow a positive setting of the *pro*-drop parameter (cf. Jaeggli & Safir 1989: introduction). For instance, Mainland Scandinavian languages are negatively set for the *pro*-drop parameter, in spite of the absence of person and number agreement (Platzack, 1987). Like Chinese, Gun-gbe lacks overt agreement and tense inflection and still it does not allow nonovert subjects (Aboh, 1998).

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

2.2 *The exception: an antecedentless empty category*

Haegeman (1997) assumes that null subjects in the relevant spoken and written registers are restricted to root clauses, hence our label RNS. She concludes that the RNS in (1b,c) does not fit any of the empty categories postulated (see (2)). As the RNS is somehow 'exceptional', she tries to account for it by adding a marked property to the grammar in which they are attested. The major ingredient of the analysis in Haegeman (1997, forthcoming) is Rizzi's (1994) formulation of the identification condition for empty categories.[7] Here, we omit all technical details of the analysis; these are spelt out in the earlier paper.

The RNS in (1b–c) is the subject of a finite clause. In this respect it is like an A′-trace (2ii) : it is an empty subject of a finite clause and it alternates with an overt constituent. The RNS differs from the A′-trace in that it seems to be antecedentless, not being related to a moved *wh*-phrase (2iia) or topic (2iib). The RNS in (1b,c) is not identified by a constituent in the clause but it picks up a referent from the discourse. Recall that we assume that the content of empty categories must be recovered. Empty categories are subject to an identification requirement. The content of an empty subject is identified by a c-commanding antecedent, i.e. an antecedent which precedes the empty category and is hierarchically higher in the structure. For instance, in (6), repeated here as (17), the A′-traces in the embedded clauses are identified by their antecedents, *who* and *Nelson* respectively.

(17) (a) Who_i do you think [$_{CP}$ [$_{IP}$ t_i will get here first]]
(b) Nelson_i , I don't think [$_{CP}$ [$_{IP}$ t_i will get here first]]

Unlike in Italian, the content of an empty subject of an English finite clause cannot be recovered by the rich verbal inflection, English inflection being notably poor. In the absence of an overt antecedent and rich inflection, the RNS should systematically violate the identification condition on empty categories.

Haegeman (1997) follows Rizzi (1994) in formulating an exception clause to the identification condition on empty categories. The essence of this proposal is that when an empty category occupies the highest, i.e. leftmost, position of the clause, then, precisely by virtue of its position, it may remain antecedentless. This is so because there simply exists no higher position in the clause. In the highest position in the clause, an empty category can be identified directly from the discourse.

The prediction from this analysis is that an English null subject in a finite clause (as in (1b–c)) may only remain antecedentless when not preceded by any syntactic position which could host an antecedent. Hence, such an antecedentless null subject should be banned from embedded clauses. If a finite embedded clause did contain a null subject, then the availability of the matrix clause inevitably would render available positions which are higher than that empty category and which may host

[7] As shown in Haegeman, Ihsane, & Koutsandréou (1998), Thrasher's (1977) erosion analysis is a precursor of Rizzi's (1994) analysis.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

antecedents. The nonovert subject of a finite clause is an A′-trace, and has a *wh*-phrase (17a) or a topic (17b) as its antecedent. [8]

In her discussion of nonovert subjects in spoken English, Schmerling (1973: 583) states that 'the English deletions occur only in the highest sentences'. Similarly, Thrasher (1977: 29) considers (18) ungrammatical. We return to the spoken register in section 6.

(18) *He thinks that [ec] will help you.

As mentioned, nonovert subjects of finite clauses are like A′-traces (2ii). The antecedent of the nonovert subject in (18) would have to be a sentence-initial interrogative pronoun or a topic . No such antecedent is available. A′-traces cannot take the subject of the higher clause as an antecedent (cf. (6)).

Haegeman (1997: 230: note 6) did signal one problematic example from diary writing, namely (19), in which the subject of a *since*-clause is nonovert. The example was taken to be an isolated case and was left unanalysed.

(19) Ten weeks *since left home* [italics ours][9]

In this paper, we reconsider the status of examples such as (1d–g), (18) and (19) in the relevant written register. We will examine the distribution of the overt and nonovert subjects in relation to clause type in English diary extracts.

A further prediction of Haegeman's account (1997, forthcoming) is that null subjects are banned from root clauses in which a *wh*-constituent precedes the subject. The nonovert subject is signalled by [ec] (for empty category).

(20) (a) *When will [ec] see him again?
(b) *What a nice present [ec] was!

In (20a) and in (20b) the preposed *wh*-phrases occupy a sentence-initial position, [Spec,CP]. This position could have hosted an antecedent for the null subject (i.e. a *wh*-phrase or a topic). (20a) and (20b) violate the identification condition on empty categories. The nonovert subject could have been identified by an antecedent, and it is not.

In (20c), the inverted auxiliary in the functional head C^o, to the left of the subject, will lead to the projection of CP, a projection headed by C^o. Once CP is projected, its specifier-position becomes available and could host an antecedent for the null subject.

(20) (c) *$[_{CP}$ $[_{C^o}$Are] $[_{IP}$ [ec] going to lunch]]?

Summing up, the earlier account assumes that a subject of a finite clause can remain nonovert just in case there is no position for a potential antecedent. In such contexts, and only in such contexts, the identification condition for empty categories

[8] See also Haegeman (1997), who assumes that root clauses may lack the layer CP which dominates IP.

[9] From the hand-written notes of Michael Palin's diary entry for 5 Dec. 1988 (cover of Dutch edition: *Around the world in eight days*. BBC books). Thanks to Frank Nuyts (p.c. 14 Aug. 1995) for signalling this example.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

is irrelevant. The null subjects in (1b,c) are empty categories which are licensed by virtue of the non-availability of an antecedent position. Their content is identified directly via the discourse.

3 Survey of the data

3.1 Root null subjects

We have examined the distribution of null subjects in the following diary extracts.

(i) E. Smart, *Necessary secrets*. 1991: 163–82
(ii) Anthony Powell, *Journals 1982–1986.* 1995: 77–96
(iii) Sylvia Townsend Warner, *Diaries*. 1995: 213–32
(iv) Sylvia Plath, *Journals*. 1995: 284–303
(v) Kenneth Williams, *Diaries*. 1994: 1–22

We have also examined extracts from two fictional diaries.

(vi) Alison Weir, *Does my bum look big in this?* 1997: 107–26
(vii) Sue Townsend, *Adrian Mole: the wilderness years.* 1993: 3–22, 75–94, 130–49

Table 1 is a survey of the distribution of null subjects for these extracts. We distinguish three types of clauses. For each extract, the first line groups root clauses with *wh*-preposing (both interrogative and exclamative) and with subject-auxiliary inversion (typically *yes*/*no*-questions); the second line presents our findings for declarative root clauses; the third line those for embedded clauses. For each clause type, we give the total number of sentences with overt subjects ('overt SU'), the total number of sentences with nonovert subjects ('null SU'), the total number of finite clauses ('total finite S'), and the percentage of sentences with null subjects per clause type ('% null SU').

The findings in these extracts confirm the predictions made by Haegeman (1997): subject omission is a root phenomenon. Apart from one example in the Williams diary, to which we return presently, we found no ENS.

The frequency of root-subject omissions varies across authors, remaining as low as 2 per cent in the Smart extract, and rising to almost 25 per cent in Plath. In all extracts, the RNS is strictly confined to declaratives. Sentences with *wh*-preposing or inversion do not have null subjects.

Our findings are supported by Ihsane's (1998) study of thirty pages of Virginia Woolf's diary (entries for 1940) summarized in table 2.

Table 3 is based on a short diary-style column, entitled 'A life in the day of William Boyd' in *The Sunday Times Magazine* (8 Feb. 1998: 58).[10] The null subject is very frequent, but it remains a root phenomenon.

[10] Thanks to Margaret Clayton for bringing this text to our attention.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Table 1 *Null subjects in diary extracts*

Extract		overt SU	null SU	total finite S	%null SU
Smart:163–82	*wh+y/n* root	30	0	30	0
	decl. root	472	10	482	2.07
	embedded	156	0	156	0
Powell:77–96	*wh+y/n* root	0	0	0	0
	decl. root	415	18	433	4.16
	embedded	210	0	210	0
Warner:213–33	*wh+y/n* root	12	0	12	0
	decl. root	354	15	369	4.07
	embedded	179	0	179	0
Plath:284–303	*wh+y/n* root	49	0	49	0
	decl. root	296	98	394	24.87
	embedded	154	0	154	0
Williams:1–22	*wh+y/n* root	12	0	12	0
	decl. root	134	195	329	59.27
	embedded	64	1?	65	2?
Weir: 107–26	*wh+y/n* root	22	0	22	0.00
	decl. root	311	19	330	5.7
	embedded	251	0	251	0.00
Townsend:3–22	*wh+y/n* root	12	0	12	0.00
	decl. root	249	8	257	3.11
	embedded	91	0	91	0.00
Townsend: 75–94	*wh+y/n* root	6	0	6	0.00
	decl. root	205	43	248	17.34
	embedded	70	0	70	0.00
Townsend: 130–49	*wh+y/n* root	3	0	3	0.00
	decl. root	229	8	237	3.38
	embedded	73	0	73	0.00

Table 2 *Woolf*, Diary *1985: 310–29 (see Ihsane, 1998)*

	overt SU	null SU	total finite S	% null SU
wh+y/n root	51	0	51	0.00
decl. root	827	111	938	11.83
embedded	287	0	287	0.00

Table 3 *Boyd* (Sunday Times)

	overt SU	null SU	total finite S	% null SU
wh+y/n root	1	0	1	0.00
decl. root	35	40	75	53.33
embedded	19	0	19	0.00

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

3.2 Embedded null subjects (ENS)

While in line with our expectations, the Williams extract does contain one ENS:

(21) (a) Gay evening in S'pore restaurant – so many gins [ec] wasn't true! (Williams, *Diaries*: 12)

we interpreted this example as containing an embedded clause:

(21) (b) [We had] So many gins [that it] wasn't true

On two other occasions in the extract, Williams uses the construction with an overt subject in the result clause (italics ours).

(21) (c) he and I went into the village – so gay *it* wasn't true (Williams, *Diaries*: 9)
(d) Geoff and I went into Singapore. So gay *it* wasn't true (Williams, *Diaries*: 13)

While the first group of extracts confirms our expectations, an extension of the material has revealed that the ENS is attested. The following extracts were examined.

(i) Allen Ginsberg, *Journals 1954–1958*. 1995: 73–92; 95–115
(ii) Helen Fielding, *Bridget Jones's Diary*. 1996: 7–26; 139–58; 204–23

Table 4 presents the results of the analysis.

Ginsberg's diary shows a discrepancy between the frequency of RNS and that of ENS. There are around 33 per cent RNS, only 3 to 4 per cent ENS. In Fielding's fictional diary, ENS reaches higher percentages throughout. In the third extract, ENS is almost as frequent as RNS. (22) and (23) illustrate ENS in Ginsberg and in Fielding, respectively.

(22) (a) Wonder if [ec] ever make it so? (Ginsberg: 77)
(b) also said [ec] should make girls (Ginsberg: 99)
(c) sexual ecstasy as I thot [sic] [ec] yet have (Ginsberg: 81)
(d) next time [ec] see him must be alive (Ginsberg: 77)
(e) The most important price of introspection – if [ec] can only rid self-conscious (Ginsberg: 82)
(f) When [ec] saw him at noon, he'd been in North Beach all last night (Ginsberg: 92)
(g) when [ec] only wanted fuck for own sake (Ginsberg: 100)
(h) Then I resolved to stay night here, though with money [ec] might have gone to 755 Pine (Ginsberg: 78)
(i) as [ec] also must remember seeing him with Peter recently (Ginsberg: 100)
(j) looked face-fallen a moment as [ec] left (Ginsberg: 78)
(k) Deceiving her impossible as [ec] would be unsuccessful in long run (Ginsberg: 49–50)

(23) (a) was worried that [ec] might split (Fielding: 227)
(b) Think [ec] will cross that last bit out as [ec] contains mild accusation (Fielding: 25)
(c) cannot believe [ec] have not realised this before (Fielding: 156)
(d) Start to wonder whether [ec] am really good friend (Fielding: 261)
(e) Understand where [ec] have been going wrong (Fielding: 97)

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Table 4 *Null subjects in Ginsberg and Fielding*

Extract		overt SU	null SU	total finite S	% null SU
Ginsberg: 73–92	*wh+y/n* root	34	0	34	0
	decl. root	301	150	451	33.26
	embedded	162	7	169	4.14
Ginsberg: 85–119	*wh+y/n* root	21	0	21	0
	decl. root	258	126	384	32.81
	embedded	101	3	104	2.88
Fielding: 7–26	*wh+y/n* root	5	0	5	0.00
	decl. root	146	29	175	16.57
	embedded	107	15	122	12.30
Fielding: 139–58	*wh+y/n* root	8	0	8	0.00
	decl. root	137	55	192	28.65
	embedded	133	17	150	11.33
Fielding: 204–23	*wh+y/n* root	3	0	3	0.00
	decl. root	154	19	173	10.98
	embedded	77	9	86	10.47

(f) Give all clothes which [ec] have not worn for two years or more to homeless (Fielding: 3)
(g) But even that is inadvisable since [ec] am fat (Fielding: 17)
(h) Delighted by, well, anything – as [ec] always am if [ec] is not work – (Fielding: 22)
(i) stopped when [ec] got to it (Fielding: 130)
(j) suddenly start thinking of former boyfriend Peter with whom [ec] had functional relationship for seven years, until [ec] finished for heart-felt agonizing reason [ec] can no longer remember (Fielding: 190)
(k) but only string [ec] have got is blue (Fielding: 266)
(Koutsandréou, 1998: appendix; Ihsane, 1998: appendix)

ENS is found in different types of embedded clauses: complement clauses with (23a) or without *that* (22b–c), interrogative complement clauses (23d–e), relative clauses (22d, 23f,j,k), and adjunct clauses (22e–k, 23f–k).

We found no subject omission in the fifty-five root sentences with *wh*-preposing, or with subject-auxiliary inversion.

Table 5 summarizes the distribution of overt and of nonovert subjects in eight recent diary extracts which appeared as newspaper columns. (v)–(vii) echo *Bridget Jones's diary*.

(i) A 1,000-word instalment of *Bridget Jones's diary*, in *The Independent* (24 Dec. 1996) (this extract is not included in the book)

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Table 5 *Diary columns*

		overt SU	null SU	total finite S	% null SU
Fielding: *Independent*	*wh+y/n*	0	0	0	0.00
	decl. root	15	32	47	68.09
	embedded	14	8	22	36.36
Fielding: *Daily Teleg.*	*wh+y/n*	0	0	0	0.00
	decl. root	48	4	52	7.69
	embedded	8	3	11	27.27
Fielding: *Newsweek*	*wh+y/n*	0	0	0	0.00
	decl. root	15	38	53	71.70
	embedded	10	17	27	62.96
Faulks	*wh+y/n*	0	1	1	100.00
	decl. root	30	27	57	47.37
	embedded	5	3	8	37.50
Barham	*wh+y/n*	2	0	2	0.00
	decl. root	3	35	38	92.11
	embedded	25	4	29	13.79
Bywater	*wh+y/n*				
	decl. root	24	47	71	66.20
	embedded	17	9	26	34.62
May	*wh+y/n*	1	0	1	0.00
	decl. root	14	22	36	61.11
	embedded	10	7	17	41.18
Clinton	*wh+y/n*	0	0	0	0.00
	decl. root	21	22	43	51.16
	embedded	13	1	14	7.14

(ii) A 1,000-word instalment of *Bridget Jones's diary*, in *The Daily Telegraph* (15 Aug. 1998: 18)
(iii) A 2,000-word fictional diary, 'Helen Fielding's diary', by Helen Fielding, in *Newsweek* (29 June 1998: 54–5)[11]
(iv) 'Jack the Lad', a 1,000-word text by Sebastian Faulks, in *The Guardian* (7 Jan. 1998: 9)
(v) 'Paula Jones's diary', a 1,000-word text by Debbie Barham, in *The Independent* (20 Jan. 1998: 15)
(vi) A 1,500-word entry of the fictional *Mark Darcy's diary* by Michael Bywater in *The Observer* (7 June 1998: 28)
(vii) 'Bad Trip, the diary of Emma D. May', a 1,000-word fictional diary entry, in *The Independent on Sunday* (21 June 1998: 11)
(viii) 'Mrs Clinton's diary', a 1,000-word fictional diary entry, in *The Observer* (23 Aug. 1998)[12]

[11] Thanks to Tatiana Koutsandréou for bringing this text to our attention.
[12] Thanks to Siobhan Cottell for bringing this text to our attention.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(24) illustrates ENS from these columns.

(24) (a) Felt [ec] had to show face at post-finals party in Gary Brigg's room (Faulks)
(b) Wonder if [ec] can poss. also sue him for contravening UN weapons inspection (Barham)
(c) Cheered when [ec] telephone to report cinema usherette to manager (Faulks)
(d) Dashed down when [ec] heard policemen (Faulks)
(e) will deffo give them up if [ec] win $2m in out-of-court settlement (Barham)
(f) Although [ec] would not obviously say that in court because might prejudice against me (Barham)
(g) B. said [ec] were planning on shopping locally (Bywater)
(h) if [ec] want family life with someone, will have to be caring (Bywater)
(i) Will miss him, though [ec] never understood why we all had to have sex with him (Emma D. May)
(j) Trapped in own personal Groundhog Day hell, where [ec] keep waking up in airplane toilet with profound sense of déjà vu. (Emma D. May)
(k) can only imagine expression when [ec] spotted ice-pick lying next to sofa (Clinton)

ENS is not exclusive to the fictional diary, as shown by the data from Ginsberg (22), and – more restrictedly – from Palin (19) and Williams (21). A fictional diary does not necessarily contain ENS, either, as shown by data from Weir and Townsend. The contrast between Weir (cf. table 1) and Fielding (cf. table 4) is particularly interesting since they were published in the same period and the content and tone of the novels are closely similar.

One factor that probably plays a role in determining the high rate of subject drop in the diary as a newspaper column (including the Bridget Jones diary, which was originally a newspaper column) is the editorial constraints imposed on length. For example, the instalments of *Bridget Jones's diary* had to be kept to 1,000 words.[13] The ENS is not restricted to the diary context, as suggested by the following examples, in which (25a) illustrates the phenomenon as a postcard text, and (25b–c) illustrate the phenomenon in notes:

(25) (a) But postcards still have one advantage. … Thus, to your boyfriend, for example, you might write: 'Having a lovely time. Glad you're not here, *as [ec] have decided to dump you*. …' (*Sunday Times, Style magazine* (12 Oct. 1997: 12) (italics ours)
(b) Mustapha Merzouk, 35, manager of Joe's Café. …
Wears Joseph suits for work. Likes 70's style, would buy them, though wouldn't have picked this outfit. (Tamsin Blanchard, 'Wide Boys wear the trousers', *Independent*. 14 April 1998)
(c) What did was take 'sha' and … (Janda, 1985: 445, Horsey, 1998: 53)

Barton (1998: 54) offers (25d) as an example of telegraphic style:

(25) (d) Regret cannot attend.

[13] Thanks to Helen Fielding for pointing this out to us.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

4 Discussion

4.1 The data

The ENS data presented in section 3.2 are problematic for accounts in which the RNS is tied to the left periphery of the clause (Haegeman, 1997; Horsey, 1998). In the following sections we describe the distribution and interpretation of the ENS in the relevant written register.

4.2 Clause types

As discussed in section 3.2, the ENS appears in a wide range of clauses, including declarative complement clauses with and without *that* ((23a) and (23b), respectively), embedded interrogative clauses (22a, 23d,e), adjunct clauses (22f–k, 23b, g, h), and relative clauses with overt (23f, j) or nonovert relativizer (22d, 23j, k).

4.3 The interpretation of the ENS

4.3.1 Grammatical features

Ihsane (1998) shows that, like the RNS in the diary register, the ENS is not restricted to the first person singular. The examples in (26) illustrate the interpretations of the ENS. Second person ENS, such as those in (26b), are rare.[14]

(26) Interpretation of ENS

	person/number	example
(a)	1sg	but even that is inadvisable since [ec] am fat (Fielding: 17)
(b)	2sg	Trouble is, Rebecca's stings are aimed so subtly at one's Achilles' heels, like Gulf War missiles going 'Fzzz Whoosssh' through Baghdad hotel corridors, that [ec] never see them coming (Fielding: 146)
(c)	3sg	Think [ec] will cross that bit out as [ec] contains mild accusation (Fielding: 25)
(d)	1pl	[ec] Could all go out for meal (Fielding: 78)
(e)	3pl	2 Bloody Mary (count as food as [ec] contain Worcester sauce and tomatoes) (Fielding: 7)

The third-person ENS may have a [+human] reading and a [-human] reading (26c,e).

4.3.2 Identification of the null subject by antecedent

The identification of the ENS takes place via coindexation with a nonovert (27a–c) or overt (27d–g) antecedent in the matrix clause, or via coindexation with a salient discourse antecedent (28).

[14] We take a second-person reading to be an option here, though maybe not the only one. Note that the second-person ENS would have a generic interpretation. See also Haegeman, Ihsane & Koutsandréou (1998).

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(27) (a) [ec_1] think [ec_1] will cross that last bit out as [ec] contains mild accusation (Fielding: 25)
(b) [ec_1] wish [ec_1] was dead (Fielding: 68)
(c) in future [ec_1] must get straight up when [ec_1] wake up (Fielding: 93)
(d) I_1 will give all clothes which [ec_1] have not worn for two years (Fielding: 3)
(e) when $Sharon_1$ rang saying [ec_1] wouldn't swear on her (Fielding: 261)
(f) Then I_1 resolved to stay night here, though with money [ec_1] might have gone to 755 Pine (Ginsberg: 78)
(g) Actually, if [ec_1] don't do something about it I_1'll start picking at it and next thing I'll have no fingernail left (Fielding: 144)
(h) He_1 gave me a look when [ec_1] walked past (Fielding: 24)

(28) (a) even that is inadvisable since [ec] am fat (Fielding: 17)
(b) Trouble is, Marco says [ec] am supposed to (Fielding: 266)
(c) Delighted by, well, anything – as [ec_1] always am if [ec_2] is not work (Fielding: 22)
(d) This is what [ec] sent (Fielding: 23)
(e) heart gave great lurch when [ec] got back from loos (Fielding: 44)
(f) Party got off to a bad start when [ec] could not see anyone (Fielding: 98)
(g) Then [ec_1] started ringing again when [ec_2] went away (Fielding: 130)
(h) [ec_1] Giddily reflect that America really is Land of the Free in sense that [ec_2] do not have layers of tradition, class and propriety giving fear of being somehow wrong all the time. (Fielding, *Newsweek*: 55)[15]

Sometimes, the grammatical features of the ENS are partly or fully recoverable from the verb inflection. In (27a), third-person-singular *contains* allows identification of the subject, so does first-person-singular *am* in (28a, c). In (27b), *was* is either the first person singular or the third person singular. In other examples, the grammatical features of the ENS are not (fully) recoverable from the verb inflection. This is the case for the modals *will* (27a), *wouldn't* (27e), *might* (27f), and *could* (28f), and for the lexical verbs *walked* (27h), *sent* (28d), *got* (28e), and *went* (28g).

In (28i), the third person singular ENS is an expletive in construction with an extraposed clause:

(28) (i) Think [ec] was right decision [not to attempt to lure] (Fielding: 215)

4.3.3 The choice of verb

We have not yet identified any restrictions on the type of verb associated with an ENS. In (29) the reader finds examples with lexical verbs (29a–e), perfect *have* (29f, g), progressive *be* (29h), copula *be* (29i), modals (29b, j) (italics ours):

(29) (a) Brief moment of wild joy when [ec] *discovered* bunch of roses (Fielding: 50)
(b) Think [ec_1] *will* cross that bit out as [ec_2] *contains* mild accusation (Fielding: 25)
(c) Then [ec_1] *started* ringing again when [ec_2] *went* away (Fielding: 130)[16]
(d) In future [ec_1] must get straight up when [ec_1] *wake* up (Fielding: 93)

[15] In this example the antecedent of [ec_2] is 'the inhabitants of America', made accessible, though not mentioned, in the context.

[16] As signalled by Martin Brett, the ergatives and intransitives are accompanied by further material (particles). It could be that null subjects tend to occur more regularly in longer sentences. Thanks to Martin Brett and David Adger for discussion.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(e) He gave me a look when [ec$_1$] *walked* past (Fielding: 24)
(f) Wish [ec] *had* not been born (Fielding: 78)
(g) Understand where [ec] *have* been going wrong (Fielding: 97)
(h) Realise [ec] *was* using telly remote control by mistake (Fielding: 153)
(i) Wish [ec] *was* dead (Fielding: 68)
(j) Think [ec] *would* like to move to New York (Fielding: 142)

Lexical verbs may be transitive (29a, b), ergative (29c, d), and intransitive (29e). Both present-tense forms (29b, d, g) and past-tense forms (29a, c, e, f, h, i, j) occur with an ENS.

4.4 Root clauses with wh*-preposing and inversion*

The analysis in Haegeman (1997) predicts that null subjects should be excluded from root sentences with *wh*-preposing and with subject–auxiliary inversion. We refer to the discussion in section 2.1.4. The relevant examples are repeated for the reader's convenience.

(30) (a) (Quando) tornerà [*pro*]?
(when) will-come-back [*pro*]
'(When) will he/she return?'
(b) *(When) will [ec] invite me again (Haegeman, 1997: 239)

Texts with ENS lack null subjects in root sentences with *wh*-fronting (30b). However, given the low numbers of sentences with *wh*-fronting in our extracts, the absence of the null subject may be a sampling accident.

The restricted data also do not allow us to decide whether or not the extracts in which ENS is attested allow a null subject in *yes/no*-questions with subject–auxiliary inversion. The only potential example we found is (31a), to which we might assign the interpretation in (31b).

(31) (a) Someone called Hurst scored three points. Also comes from Essex. Note to self: *could be useful*! (Faulks, *Guardian*. 7 Jan. 1998: 9)
(b) Could [ec] be useful?

However, (31b) is not the only reading possible. (31c) is an alternative:

(31) (c) [ec] could be useful, maybe.[17]

5 Speculations about the nature of the embedded null subject

The ENS seems, at this point, to defy all analysis in the *Principles and Parameters* framework. The question also arises how to deal with RNS. For texts only allowing RNS (see section 3.1), we might continue to adopt the earlier analysis. But for texts with both RNS and ENS, it is not clear that the two instantiate a different phenomenon. Indeed, their similarity in interpretation suggests they are the same phenomenon.

[17] Thanks to Bernadette Plunkett and Liz McCoy for pointing this out.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

As is the case for the RNS, none of the four 'classical' nonovert categories in (2) can characterize the ENS. We briefly go over the traditional empty category types here to show that they cannot be used to describe ENS. Haegeman's earlier analysis which treated the RNS as a marked phenomenon restricted to the left periphery of the root is obviously also inadequate for the ENS.

5.1 A-trace and PRO

The ENS occurs in finite clauses (32a–c) and it alternates with an overt subject (32d). Hence the phenomenon is incompatible with an A-trace analysis (2i) or with a PRO analysis (2iv), because both these empty categories cannot occur as the subjects of finite clauses, nor can they alternate with an overt NP.

(32) (a) as always *am* if [ec] *is* not work (Fielding: 22)
(b) Think [ec] *will* cross last bit out as [ec] *contains* . . . (Fielding: 25)
(c) Brief moment of joy when [ec] *discovered* bunch of roses (Fielding: 50)
(d) Think *I*'ll go and see Mum and Dad. (Fielding: 68)

5.2 A′-trace

Being the subject of a finite clause, the ENS has the distribution of an A′-trace (section 2.1.3). Returning to the analysis in Haegeman (1990a, 1990b) and in Bromberg & Wexler (1995), we might take the null subject in the diary style to be an A′-trace in the subject position with as its antecedent a nonovert operator. However, assuming the ENS to be the A′-trace of a fronted null topic, we would expect the ENS to be licit only in contexts in which leftward topicalization is possible. However, the ENS occurs in a wide range of contexts from which topicalization is rigorously excluded in English. (33) provides some examples: (33a) contains a declarative *that*-complement, (33b) an embedded interrogative complement, (33c) a relative clause, (33d) an adjunct clause, and (33e) and (33f) complement clauses to a noun. In each of the sentences in (33), topicalization of the embedded subject from the position signalled by [ec] will lead to ungrammaticality, as shown by (34).

(33) (a) Was worried that [ec] might split (Fielding: 227)
(b) Understand where [ec] have gone wrong (Fielding: 97)
(c) Give all clothes which [ec] have not worn for two years or more (Fielding: 3)
(d) heart gave great lurch when [ec] got back from loos (Fielding: 44)
(e) fact that [ec] was messaging ultimate boss (Fielding: 26)
(f) Catwoman suit is designed exclusively for long thin people like Jude with result that [ec] look like golliwog (*Independent*. 24 Dec. 1996)

(34) (a) *My blue skirt$_i$, I was worried [that t_i might split].
(b) *This relationship$_i$ I wonder [where t_i went wrong].
(c) *Mary$_i$ talked about all the clothes [which t_i has not worn].
(d) *Mary$_i$, I felt pleased [when t_i came back].
(e) *John$_i$, I regret [the fact that t_i was messaging my friend].

We have seen that the antecedent of an A′-trace occupies a peripheral position (it

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

is a fronted *wh*-phrase or topic), and that it cannot be a higher subject (6). But, as shown by (27), the ENS can be coreferential with a higher subject.

A final argument against a null topic analysis of the ENS is interpretive. (28i) has an expletive ENS. If the ENS were an A′-trace, its (nonovert) antecedent should be the null topic, but topicalization is incompatible with expletives (cf. section 2.1.3 for the same objection to a null operator analysis of RNS).

5.3 The nonovert pronoun

Empty categories are subject to the identification condition. Haegeman (1997, forthcoming) proposed that the RNS in the relevant register escapes this identification condition and – exceptionally – remains antecedentless precisely because of its peripheral syntactic position.

The identification condition poses a problem for the ENS: recall that although the antecedent of the ENS need not appear in the matrix clause, the ENS may have a discourse antecedent. This means that clause-internally, the ENS remains antecedentless. Because the ENS does not occupy the highest position in the clause, the exception clause elaborated by Haegeman (1997) to account for RNS[18] is inapplicable. It is thus not clear how the ENS can survive without an antecedent. Moreover, an ENS with a matrix antecedent still does not fit any of the typical empty category types. Being a subject of a finite clause, it ought to be an A′-trace, but being coreferential with the matrix subject, it cannot be an A′-trace (cf. section 2.1.3).

Interpretively, nonovert subjects in diaries, RNS and ENS, correspond to an overt pronoun:

(36) (a) [I] think [I] will cross that last bit out as [it] contains mild accusation (Fielding: 25)
(b) [I] wish [I] was dead (Fielding: 68)
(c) when Sharon rang saying [she] wouldn't swear on her (Fielding: 261)
(d) even that is inadvisable since [I] am fat (Fielding: 17)
(e) Delighted by, well, anything – as [I] always am if [it] is not work (Fielding: 22)
(f) Then [it] started ringing again when [I] went away (Fielding: 130)
(g) [I] Giddily reflect that America really is Land of the Free in sense that [they] do not have layers of tradition, class and propriety giving fear of being somehow wrong all the time. (Fielding, *Newsweek*: 55)
(h) [I] Think [it] was right decision not to attempt to lure (Fielding: 215)

One might take the ENS as evidence for parameter resetting and propose that in the relevant written register (see section 6 for spoken English), the *pro*-drop parameter is reset. This analysis might then be extended to subsume the RNS, especially so in the extracts containing ENS, since the major objection to a *pro*-drop analysis for the RNS was the absence of ENS.

[18] It was proposed that an empty category can remain without antecedent as long as there was no position for an antecedent available, restricting the empty category to the specifier of the root clause.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

In addition to the conceptual problems of the parameter resetting analysis discussed in section 2.1.4, further questions arise.

One problem is the identification of the ENS, in particular in those cases in which it lacks a matrix antecedent. Clearly, it cannot be argued that the finite verb displays the kind of rich agreement which unequivocally identifies the grammatical features of the ENS.

Moreover, sentences such as those in (37), with an overt embedded subject which is coreferential with the matrix subject, are problematic for an Italian style *pro*-drop account:

(37) (a) Might as well do it now while I think about it. (Fielding: 144)
(b) In fact think might put the Glossies bra on to match if I can find it. (Fielding: 145)

In Italian, typically a subject of an adjunct clause which is coreferential with a matrix subject will not be spelt out overtly (see (11b)).

Haegeman, Ihsane & Koutsandréou (1998) propose that the ENS is the result of a Chinese-style resetting of the *pro*-drop parameter in which null subjects are licensed by virtue of the absence of overt agreement (see (16)). In that paper, it is suggested that this resetting may be seen as a result of the relative poverty of the English verbal paradigm coupled with the reduction of speaker–hearer interaction in the diary register. There remain important problems with this analysis, though, and we will not adopt it here.

On the conceptual side, the proposal that the restricted speaker–interlocutor interaction leads to a resetting of a parameter raises the question of the distribution of labour between syntax properly speaking and discourse-related phenomena, two domains which are considered impermeable in the *Principles and Parameters* approach.

If there is a genuine resetting of the *pro*-drop parameter in this register, then one would expect null subjects also to be possible in inversion contexts, i.e. in root *yes/no* questions and in root *wh*-questions. Unfortunately, the relevant data are too rare to allow any conclusions.

If the resetting of the *pro*-drop parameter is linked to the poverty of English inflection, the prediction would be that ENS will not turn up in French, a language with a richer inflectional paradigm. This prediction is not borne out. If Stroumza's (1998) translation of *Bridget Jones's diary* is anything to go by, French diary writing also may allow the ENS, though probably more marginally. Out of the 127 instances of ENS in the English text, ten have been translated with an ENS. (38) contains some examples.[19]

(38) (a) Me rappelle tout à coup que portais jupe noire Lycra dernière fois (1998: 113)
remember suddenly that wore skirt black Lycra last time
(b) En cherchant le lait me suis aperçue qu'ai laissé filet (1998: 104–5)
looking for milk noticed that have left basket

[19] Sylvie Ferioli points out that such constructions could also be found in note-taking.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(c) Suis tellement énervée que me suis assise sur la télécommande (1998: 182)
am so nervous that me am seated on remote control
(d) parce que se sentait obligé de (1998: 182)
because himself felt obliged to

Another prediction of the account is that embedded null subjects should be excluded from spoken English. Though the data are not clear, this prediction may also be false. We refer to section 6 for some discussion.

5.4 Generalized pronoun ellipsis and the subject/object asymmetry

An alternative approach to subject omission in registers with both ENS and RNS would be to account for it in terms of a pronoun ellipsis rule which (optionally) operates on the output of the grammar. By this account, speakers are not claimed to have two grammars, one set positively and one set negatively for the *pro*-drop parameter. Rather, they have one grammar with an optional ellipsis.

As a first approximation, it could be proposed that economy considerations lead to a generalized ellipsis of pronouns whose content can be identified via the discourse.

If a pronoun ellipsis account is unqualified, one would expect object omission in the relevant register, and indeed some object omissions are attested, as shown in (39):

(39) (a) Today Robt. comes by at 1 to borrow dime, I haven't got, returns at 3 (Ginsberg: 80)
(b) Toy with idea of flirting energetically with anyone I think might be induced to send me one, but dismiss as immoral. (Fielding: 42; Koutsandréou, 1998: 41)
(c) Assemble shepherds pie and place in oven (Fielding: 83)
(d) Get second one and find hole on back of leg. Throw away. (Fielding: 92)
(e) Search for hairbrush. Locate in handbag (Fielding: 92)
(f) Suddenly remember had handbag in bedroom when looking for hairbrush but cannot find. (Fielding: 93)
(g) Eventually locate under clothes from wardrobe (Fielding: 93)
(h) Have only visited briefly once when 25 (Fielding, *Newsweek*. 29 June 1998: 54)
(i) Audience warm-heartedly forgives (Fielding, *Newsweek*. 29 June 1998: 55)
(j) 'Silverskin onion?' I encouraged. (Fielding: 16)

In (39k) the indirect object of *giving* – 'them' – is nonovert:

(k) Giddily reflect that America really is Land of the Free in sense that do not have layers of tradition, class and propriety giving fear of being somehow wrong all the time. (Fielding, *Newsweek*. 29 June 1998: 553)

However, generalized pronoun ellipsis seems in conflict with the subject/object asymmetry found in pronoun omission. Table 6 gives some information on object omission in the *Bridget Jones's diary* passages. The label 'object' covers pronouns and NPs functioning as direct object, indirect object, subject of a nonfinite clause, and subject of a verbless clause. We distinguish overt NP objects (second column) from overt pronominal objects (third column). Among nonovert objects, we distinguish sentences in which a potentially transitive verb lacks an object resulting

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Table 6 *Object omission in* Bridget Jones's diary

page	overt NP object	overt pronoun object	null object grammatical	null object ungrammatical
7–26	100	50	13	1? (=39j)
139–58	134	27	8	0
204–23	97	22	4	0

in a grammatical construction (fourth column) as in (40), from sentences like those in (39), where object omission has led to ungrammaticality in the standard grammar (fifth column).

(40) (a) I watched resignedly (Fielding: 10)
(b) Was trying to work on cv without Perpetua noticing (Fielding: 22)

Overall, object ellipsis is rare. It is not the case that whenever an object could have been omitted licitly (as in (40)), this is the norm in the relevant register. In (41) the verb *ask* has an overt pronominal object *him*, which is fully retrievable from the discourse and whose deletion would not have given rise to ungrammaticality.

(41) her boyfriend ... had chucked her for asking him if he wanted to come ... (Fielding: 19)

Object omission leading to an ungrammatical sentence is also unusual. In (42) the object of the verb *erase* is realized by the pronoun *it*, even though it is retrievable from the context:

(42) Maybe he has not got it yet. If only could get message back. Think will go for walk and see if can somehow go into Daniel's office and erase *it*. (Fielding: 24, italics ours)

In (43) the pronoun *it* is in construction with the extraposed clause. In spite of its straightforward clause-internal identification, the pronoun remains overt.

(43) but I didn't want him to make *it* perfectly obvious to everyone that he didn't want to (Fielding: 16, italics ours).

5.5 Pronoun ellipsis in a specifier-head configuration with AGR

That a pronoun ellipsis analysis is, however, on the right track is suggested by the reduction of the reflexive *myself* to *self* in the diary register. (44) gives some examples. The reduced form alternates with the full reflexive.

(44) (a) Note to *self* (Faulks, *Guardian*. 7 Jan. 1998: 9)
(b) Decided to have cappuccino and chocolate croissants on way to work to cheer *self* up (Fielding: 50)
(c) Feeling v. pleased with *self* (Fielding: 77)

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(d) Find *self* constantly scanning face in mirror (Fielding: 78)
(e) Have cigarette to cheer *self* up (Fielding: 92)
(f) Am starting to get carried away with idea of *self* as Calvin Klein-style mother (Fielding: 116)
(g) trying to whip up frenzy of emotion in *self* about end of war (Fielding: 121)
(h) threw *self* across (Fielding: 130)
(i) Was believing could totally reinvent *self* (Fielding: 184)
(j) as if I am imaginary person employed by *self* (Fielding: 219)
(k) Found *self* horrible centre of attention at morning meeting today (Fielding: 239)
(l) Peter depressed & unhappy with *self* (Ginsberg: 86)
(m) Flatten *self* against wall (Fielding, *Newsweek*. 29 June 1998: 55)
(n) Clearly necessary to distract *self* from nerves (Fielding, *Newsweek*. 29 June 1998: 54)
(o) peers at *self* in mirror (Bywater: 28)
(p) Hope that new image makes *self* look sultry (Barham: 15)
(q) as now in danger of crapping *self* in sleep like yucky five-year-old (Emma D. May: 11)
(r) brace *self* for touchdown (Emma D. May: 11)

In table 7 we examine the diary extracts with respect to reflexive reduction.

The pattern that emerges, at this point, is that in some diaries, only RNS is attested, while in others both RNS and ENS are found. The latter group also displays reflexive reduction. Object omission remains extremely rare.

If the coincidence between the occurrence of ENS and reflexive reduction is confirmed, both types of ellipsis could be seen as the instantiation of pronoun ellipsis licensed by a specifier–head relation with a head with abstract or overt agreement (AGR) features. The subject of the finite clause agrees with the verbal inflectional head; the pronominal part of the reflexive agrees (in number) with the reflexive component -*self*.

The formal licensing of the particular ellipsis is independent from the 'richness' of the licenser. Identification of the ellipted pronoun, i.e. recovery of its referential content, is achieved through an antecedent, either a c-commanding antecedent or a discourse antecedent.

If embedded null subjects can be licensed via the specifier–head relation then, obviously, this analysis could also be applied to the root null subjects, particularly in texts exhibiting both ENS *and* RNS. For RNS in texts which do *not* exhibit ENS, the observed restrictions on the distribution of the null subject lead us to propose provisionally that the earlier analysis (Haegeman, 1997) be maintained, at least in part. We propose that empty categories are *licensed* by AGR and that their distribution may differ because of the constraints on their *identification*.

We assume that subject ellipsis in English is formally licensed by the specifier–head relation with an agreeing head. An ellipted subject cannot be *identified* by AGR, because English AGR is 'poor'. In this sense, the English ellipted subject is different from the Italian null subject which is both licensed and identified by 'rich' agreement.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Table 7 *Pronoun deletion in reflexives*

Group 1	X+*self*	0+*self*	total	% 0+self
Plath	24	0	24	0.00
Warner	11	0	11	0.00
Smart	8	0	8	0.00
Williams	4	0	4	0.00
Weir	10	0	10	0.00
Townsend: 3–22	3	0	3	0.00
Townsend: 75–94	2	0	2	0.00
Townsend: 130–49	6	0	6	0.00
Boyd	2	0	2	0.00
Woolf: 310–29	11	0	11	0.00
Total group 1	81	0	81	0.00
Group 2	X+self	0+self	total	% 0+self
Ginsberg: 73–92	12	6	18	33.33
Ginsberg: 95–115	4	5	9	55.56
Fielding: 7–26	6	1	7	14.29
Fielding: 139–58	5	0	5	0.00
Fielding: 204–23	4	1	5	20.00
Fielding, *Independent*	0	1	1	100.00
Fielding, *Newsweek*	1	4	5	20.00
Faulks	0	1	1	100.00
Barham	0	1	1	100.00
Bywater	3	8	11	72.73
May	0	2	2	100.00
Total group 2	35	30	65	46.15

Identification of the ellipted subject is obligatory, though, and in English it can only be achieved by means of an antecedent. The difference between the distribution of the empty categories in the texts rests on the constraints on identification, i.e. on the choice of antecedent.

For texts displaying RNS only, the choice of antecedent is restricted. As long as there is an appropriate antecedent position available, this must host the identifier. If there is no antecedent position, the content of the empty category can be identified by a discourse antecedent.[20] This restricts the null pronoun to the left periphery, as before.

[20] One might assume that the subject position is left peripheral because all higher root projections (CP) are absent. Alternatively, we might postulate a discourse operator in the CP domain to which the null subject must be linked for the recovery of its antecedent. In registers with only RNS identification of the content of the null pronoun would be strictly local in that no potential antecedent may intervene between the discourse operator and the null pronoun.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

For texts exhibiting RNS and ENS, the constraints on identification of the ellipted subject are relaxed in two ways. First, even when a potential antecedent intervenes between the ellipted subject and the discourse, the null subject remains accessible for identification by a discourse antecedent. Secondly, identification may be achieved either by an antecedent in a matrix clause or by a discourse antecedent.[21] It is not clear to us at this point which (if any) grammatical properties of the null subject are responsible for differences in and the constraints on antecedent selection, and we hope to explore the issue further in future work.

6 Null subjects in spoken English

The occurrence of RNS in spoken English is reported in work by Schmerling (1973) and by Thrasher (1977). Both authors assume that ENS is ungrammatical. If we adopt the analysis outlined in the preceding section, we can reinterpret this restriction to mean that again subject ellipsis is licensed by a specifier–head relation with agreement and that in the unmarked context the null pronoun must be left-peripheral. The question arises if, as in some written styles, the condition on the identification of the null pronoun can be relaxed in spoken English.

It is certainly the case that ENS is extremely rare. If *I think* in the following passage is not taken as a parenthetical, then (45a) illustrates a case from a dialogue in a novel.

(45) (a) You know it is funny. I never knew Rosemary's last name either. She went back to Jamaica, I think lives in the country.
(Elmore Leonard, *Jackie Brown* (first published as *Rum Punch* 1992). Dell Books, p. 60)

Lawrence (1996) studies subject omission in the natural production of one young male British speaker. Her work is based on three recordings with different levels of formality: one is informal, one intermediate, and one formal. In the 2,189 relevant finite root clauses[22] she found 92 null subjects (4 per cent); in the 787 embedded clauses she found only 6 null subjects (i.e. under 1 per cent). (45b–e) are examples of ENS.

[21] Haegeman (1997) also discusses the contrast in (i):

(i) (a) Last week, met my brother
(b) *My brother, met last week

Texts displaying only RNS contain examples of subject omission with preposed adjuncts (ia), while disallowing subject omission with preposed arguments (ib). Haegeman (1997) offers an account for this contrast.

If, as we claim, the constraints on identification are relaxed in the texts with both ENS and RNS, we predict that argument preposing should not hinder the identification of the null subject. The data available do not allow us to test this prediction.

[22] We have taken out clauses with auxiliary deletion, to have data which are like those in the main body of the text. Root sentences include coordinated sentences, though, which are not included in our data.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

(45) (b) So it turned out me dad over a course of something like fifteen years or so, until went and got married, was owed upwards of about a thousand pounds. (Intermediate, lines 104–5, transcript, p. 63)
(c) they had to sell Moor Lane, the cottage – bungalow at Moor Lane cos couldn't afford it (Intermediate, line 205, transcript, p. 65)
(d) But as I say I was surprised to find that we'd lived there for six months before went to Windmill Rise (Intermediate, line 223, transcript, p. 65)
(e) he'd built the first wall flush up to the wall of the house alongside when should have had a gap. (Intermediate, lines 261–3, transcript, p. 66)

The extreme rarity of ENS (under 1 per cent) in this spoken material could tempt us to discard these data as mere performance errors. However, it may be worth pointing out that the examples are found in the recording classified as 'intermediate' and that this recording also displays the highest proportion of RNS. This might suggest that the occurrence of ENS somehow does correlate to that of RNS. In future, we will try to further investigate spoken data to establish the status of examples such as those in (45).

7 Summary

This paper examines embedded null subjects (ENS) in written English. Contrary to what is predicted by the analysis in Haegeman (1997, forthcoming), the ENS is attested. Our paper offers a description of the ENS and discusses the problems raised for the theory of empty categories in the *Principles and Parameters* framework.

The ENS does not correspond to any of the traditional empty categories postulated in the generative literature. We have exploited the observation that the occurrence of ENS seems to coincide with the ellipsis of the pronominal component of the *self*-reflexive. In speculative terms, we propose that RNS, ENS and reduction of reflexives are instantiations of pronoun ellipsis which is formally licensed by the association with a head with agreement features. Tentatively we propose a unified analysis in which both RNS and ENS are formally licensed by a specifier–head relation with AGR. In registers with RNS but without ENS, the recovery of the null pronoun is subject to a strict locality condition imposing the constraint that the null pronoun be in a left-peripheral specifier. In registers with ENS, the constraints on the identification of the null pronoun are relaxed.

Authors' address:
Department of Linguistics
University of Geneva
CH-1211 Geneva 4
Switzerland
haegeman@lettres.unige.ch
ihsane3@uni2a.unige.ch

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

References

Aboh, E. (1998). From the syntax of Gun-gbe to the grammar of Gbe. PhD Diss. University of Geneva.

Barton, E. L. (1998). The grammar of telegraphic structures. *Journal of English Linguistics*, **26**: 37–67.

Bromberg, H. & K. Wexler (1995). Null subjects in *wh*-questions. In Schütze, C. T., J. B. Gander & K. Broihier (eds.), *Papers on language processing and acquisition*. *MIT Working Papers in Linguistics* **26**: 221–48.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge: MIT Press.

Haegeman, L. (1987). Complexity and literary prose: some suggestions for formalization. *Language and Style* **20**: 214–22.

Haegeman, L. (1990a). Non-overt subjects in diary contexts. In Mascaro, J. & M. Nespor (eds.), *Grammar in progress.* Foris: Dordrecht.

Haegeman, L. (1990b). Understood subjects in English diaries. *Multilingua* **9**: 157–99.

Haegeman, L. (1997). Register variation, truncation and subject omission in English and in French. *English Language and Linguistics* **1**: 233–70.

Haegeman, L. (forthcoming). Adult null subjects in non *pro*-drop languages. In Friedemann, M.-A. & L. Rizzi (eds.), *The acquisition of syntax.* London: Addison, Wesley and Longman.

Haegeman, L., T. Ihsane & T. Koutsandréou (1998). Embedded null subjects in English diaries. Unpublished ms. University of Geneva.

Horsey, R. (1998). *Null arguments in English registers. A Minimalist account*. BA thesis, La Trobe University. Australia.

Huang, J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry* **15**: 531–74.

Huang, J. (1989). *Pro* drop in Chinese: a generalized control theory. In Jaeggli, O. & K. Safir (eds.), *The null subject parameter.* Dordrecht: Kluwer. 185–214.

Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

Ihsane, T. (1998). The syntax of diaries: grammar and register variation. Mémoire de licence. University of Geneva. Department of Linguistics. Ms.

Jaeggli, O. & K. Safir (1989). *The null subject parameter*. Kluwer: Dordrecht.

Janda, R. J. (1985). Note-taking English as a simplified register. *Discourse Processes* **8**: 437–54.

Koutsandréou, T. (1998). Implicit subjects in English adult language. Petit mémoire. University of Geneva. English Department. Ms.

Lasnik, H. & T. Stowell (1991). Weakest crossover. *Linguistic Inquiry* 22: 687–720.

Lawrence, H. (1996). An examination of the effect of style on the realisation of subjects. MA dissertation. University of York.

Platzack, C. (1987). The Scandinavian languages and the null-subject parameter. *Natural Language and Linguistic Theory* **5**: 377–401.

Raposo, E. (1986). The null object in European Portuguese. In Jaeggli, O. & C. Silva-Corvalan (eds.), *Studies in Romance linguistics*. Dordrecht: Foris. 391–419.

Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* **17**: 501–58.

Rizzi, L. (1993). A parametric approach to comparative syntax: properties of the pronominal system. *English Linguistics* 10: 1–27. Also in Haegeman, L. (ed.), *The new comparative syntax.* London: Addison, Wesley and Longman. 268–85.

Rizzi, L. (1994). Early null subjects and root null subjects. In Hoekstra, T. & B. Schwartz (eds.), *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins. 151–77.

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press

Rizzi, L. (1995). Some notes on linguistic theory and language development: the case of root infinitives. *Language Acquisition* **3/4**: 371–93.
Schmerling, S. (1973). Subjectless sentences and the notion of surface structure. In Corum, C., T. C. Smith & A. Weiser (eds.), *Papers from the ninth regional meeting of the Chicago Linguistic Society*: 577–86.
Thrasher, R. (1977). *One way to say more by saying less. A study of so-called subjectless sentences.* Kwansei Gakuin University Monograph Series Vol. 11. Tokyo: The Eihosha Ltd.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* **40**: 21–81.
Wilder, C. (1994). Some properties of ellipsis in coordination. *GenGenP* **2**: 23–61.

Corpus material

Fielding, H. (1996). *Bridget Jones's diary*. London: Picador.
Ginsberg, A. (1995). *Journals 1954–1958*. Ed. G. Ball. London: Penguin Books.
Plath, S. (1983). *The journals of Sylvia Plath.* Eds. T. Hughes & F. McCollough. New York: Ballantine Books.
Powell, A. (1995). *Journals 1982–1986.* London: Heinemann.
Rendell, R. (1994). *A sleeping life. An Inspector Wexford mystery.* London: Arrow Books.
Smart, E. (1991). *Necessary secrets*. Ed. A. van Wart. London and Edinburgh: Paladin (HarperCollins).
Stroumza, A. (1998) *Le journal de Bridget Jones*. Paris: Albin Michel.
Townsend, S. (1993). *Adrian Mole: the wilderness years*. London: Mandarin Paperbacks.
Townsend Warner, S. (1995). *The diaries of Sylvia Townsend Warner*. Ed. C. Harmann. London: Virago.
Weir, A. (1997). *Does my bum look big in this?* London: Coronet Books.
Williams, K. (1994). *The Kenneth Williams diaries*. Ed. R. Davies. London: HarperCollins.
Woolf, V. (1985). *The diary of Virginia Woolf, Volume 5: 1936–1941*. Eds. A. Olivier Bell assisted by A. McNeillie. Middlesex: Penguin Books. First published by Chatto and Windus: The Hogarth Press (1984).

https://doi.org/10.1017/S1360674399000155 Published online by Cambridge University Press