Is there a little pro? Evidence from Finnish.

Anders Holmberg


**Abstract**

The traditional view of the null subject as *pro* identified by Agr (the $\emptyset$-features of I) cannot be maintained in a theory where Agr is uninterpretable. Two hypotheses are compared with regard to the predictions they make for Finnish null subject constructions: (A) Agr is interpretable in null-subject languages, and *pro* is therefore redundant; (B) Null subjects are specified but unpronounced pronouns which assign values to the uninterpretable features of Agr. Since Finnish observes the EPP and has an expletive pronoun, Hypothesis A predicts that null subjects should co-occur with expletives. The prediction is false, favouring B over A. A typology of null subjects is proposed: Null bound pronouns and null generic pronouns in partial null-subject languages, including Finnish, are D-less $\emptyset$Ps, and so are null subjects in consistent null-subject languages with Agr, such as Spanish, Greek, etc. Null 1st and 2nd person subjects in Finnish are DPs which are deleted. Null pronouns in languages without Agr, such as Chinese or Japanese are the only true instances of *pro*, a minimally specified null noun.

key words: null subject, empty category, phi-feature, expletive, pronoun,
Sentence (1) is an example of a so called null subject.

(1) Olen väsynyt. (Finnish)
    be-PRES.1SG tired
    ‘I’m tired.’

The sentence has no overt subject, yet, according to standard Principles-and-Parameters theory it has a subject which is an underspecified, phonetically empty subject pronoun, so called (little) pro, formally licensed and interpreted by virtue of the agreement on the finite verb or auxiliary.

This paper is a scrutiny of this hypothesis, which goes back to works such as Chomsky 1982 and Rizzi 1986, in the light of more recent developments in syntactic theory, particularly the feature theory of Chomsky 1995: ch. 4, and subsequent work by Chomsky and others. I will argue that there is a null (unpronounced) pronoun in the subject position in (1), but with properties which are different from those of pro in standard Principles-and-Parameters theory. I thereby disagree with claims made in Manzini & Roussou 1999, Manzini & Savoia 2002, Platzack, 2003, 2004, and with some reservations, Anagnostopoulou & Alexiadou 1998, according to which constructions such as (1) do not have a subject at any level of representation other than the nominal features realized on the verb or auxiliary. Crucial evidence in support of my hypothesis comes from null subject constructions in Finnish. This evidence leads me to conclude that there are several types of syntactically projected null subjects. One commonly occurring type is a null weak pronoun, roughly as characterized in Cardinaletti & Starke 1999: a pronoun specified for φ-features but lacking D, and therefore incapable of (co-)referring, without the help of a D-feature in I, in a manner detailed in the paper. Another type of null subject is a DP which is deleted under usual conditions of
recoverability. A third type is the classical pro, in the sense of a minimally specified nominal category, a bare, $\phi$-less noun. This type is found only in languages without unvalued $\phi$-features in I, that is without Agr. In languages with Agr, the subject must have inherently valued $\phi$-features, in order to value Agr, thus excluding pro as subject.

1. The GB theory of empty categories

Pro was introduced in Chomsky 1982 as part of a theory of empty NPs, which in turn formed part of the theory of NP-types and Binding, one of the cornerstones of the Government-Binding (GB) program. According to this theory, the different types of empty nominal categories that had been identified and whose properties had been intensely investigated, namely NP-trace, wh-trace, PRO, and now also pro, were really special cases of the same category, an empty nominal category with no inherent properties apart from (presumably) nominal categorial features and maximal X-bar level: $[\text{NP } e]$. The grammatical properties of $[\text{NP } e]$ are ‘functionally determined’, that is, they are determined by the syntactic relations that it enters into, particularly the binding relation:

(2)  $[\text{NP } e]$ is

a. wh-trace if it is locally A-bar bound,

b. NP-trace if it is locally A-bound from a non-$\theta$-position,

c. PRO if it is locally A-bound from a $\theta$-position,

d. pro if it is governed by strong enough INFL or by a clitic.

Chomsky 1982
The four types of empty categories correspond to three types of overt NP, namely anaphors, pronouns, and R-expressions, defined by the two binary features [+/-pronominal, +/-anaphor], plus one type which for principled reasons can only be empty, namely PRO. The result is a theory of almost unparalleled elegance within formal linguistic theorizing: A range of phenomena (the different empty categories), each with their distinctive properties, are seen to be special cases of a single phenomenon (the featureless empty NP), their distinctive properties derived from other independent properties, primarily the binding relation they enter into. This phenomenon is then further unified with another set of phenomena, namely different forms of overt NPs, their distinctive properties being ultimately derived from variation with regard to the value of two primitive binary features. Together with certain other axioms of the theory, especially the Theta Criterion, the Case Filter, and the Empty Category Principle, this theory could explain an impressive range of phenomena to do with the syntactic distribution of NPs, across a wide range of languages, potentially including all natural languages.

Impressive though it was, the theory of empty categories soon began to crumble. The main problems in the 80ies and early 90ies had to do with the Binding Theory, which the theory was inextricably linked with. The main problem there was the discovery of various NP types that did not fit into the restrictive framework of the classical Binding Theory, including long distance reflexives, logophoric pronouns (see Huang 2000 for an overview), SE vs. SELF type anaphors (Reinhart & Reuland 1993), and dependent pronouns (Fiengo & May 1994); see Safir 2004 for discussion.

An even greater challenge for the theory of empty categories in (2) was the emergence of the Minimalist Program. The GB theory of empty categories is incompatible with the Minimalist Program as characterized in Chomsky 1995 and subsequent works. Perhaps most strikingly this is the case for traces. In GB theory (Chomsky 1981, 1982) a moved category
and its trace are two distinct categories. They form a Chain, which means that they share a Theta-role and a Case, but they are nevertheless two distinct categories, which may belong to distinct NP-types; for example in the case of (3), *John* is an R-expression and the trace an anaphor.

(3) *John* was arrested [NP e] .

In Minimalist derivational theory *John was arrested* is derived by (a), merging the lexical item *John* with *arrest* forming the VP [arrest *John*], (b) expanding the tree by merging more categories one by one with the derived tree, (c) merging *John* a second time, to satisfy the EPP-feature of T, (d) spelling the structure out without spelling out the lower copy (or occurrence) of *John* (see Chomsky 1995, Nunes 2000). That is to say, there is no trace, in the sense of an empty category with its own distinct properties, present at any stage of the derivation. Wh-movement follows the same principles: a whP is first merged as an object, subject, or adverbial, etc., and subsequently remerged with CP, potentially several times, in the case of long-distance extraction. There are several copies (or occurrences) of the whP, but no traces/empty categories.

The two types of traces gone, all that remains of the theory of empty categories in (2) is PRO and *pro*. I will put PRO aside, returning to it briefly at the end of the paper, and consider the position of *pro* in the Minimalist Program in some detail.

2. Little *pro* and uninterpretable features

The most authoritative theory of *pro* within the GB program is the one articulated by Rizzi 1986, building on earlier work by Chomsky 1981, 1982, Rizzi 1982, Bouchard 1984, among
others: *pro* is inherently unspecified for \( \phi \)-feature values. Its distribution is regulated by a licensing condition and a recovery (or identification) condition, as follows (see Rizzi 1986):

(4) **Licensing:** *pro* is Case-marked by \( X^o_y \), where \( y \) is parametrized.

**Identification:** *pro* inherits the \( \phi \)-feature values of \( X^o_y \) (if it has \( \phi \)-features; if not, *pro* gets a default interpretation, typically *arb*).\(^1\)

In the case of null subjects, *pro* is Case-marked by I(NFL), which qualifies as a *pro*-licensing \( X^o_y \) in some languages but not others, the standard assumption being that this correlates with richness of agreement. Identification is then ensured as *pro* inherits the \( \phi \)-feature values of INFL. In (6), representing the (relevant part of) the structure of (5), *pro* is licensed because it is governed and assigned Case by I, which in Spanish qualifies as \( X^o_y \).

(5) Están cansadas.  (Spanish)

be-3PL tired-F.PL

‘They are tired.’

(6) \[\text{IP} \quad \text{[3PL]} \quad \text{I'} \quad \text{I} \quad \text{[3PL]}\]

The content of *pro* is identified as *pro* inherits the \( \phi \)-feature values of I.

The theory of *pro* outlined above cannot be maintained in a theory making the distinction between interpretable and uninterpretable features which plays a crucial role in Chomsky 1995: ch. 4 and subsequent work by Chomsky and others. Chomsky argues that there are two varieties of \( \phi \)-features: an interpretable and an uninterpretable variety. The person, number
and gender features of an NP (or DP) are interpretable, restricting the denotation of the NP. The person, number, or gender features which appear on a verb, auxiliary or adjective are uninterpretable as they do not restrict the denotation of these categories.

(7) Las chicas están cansadas.

the girls be-F.3PL tired-F.PL

(7) asserts that a group of female individuals excluding the speaker and the addressee (the denotation of the NP las chicas) each have (some degree of) the same, indivisible and genderless property of being tired (the denotation of the predicate están cansadas). The sentence does not, for example ascribe to the girls a particular female way of being tired, or, at least not necessarily, repeated occurrences of being tired.² By definition uninterpretable features cannot survive until LF, so they must be eliminated in the course of the derivation of LF. However, they may be, and typically are, visible in PF. According to Chomsky (2000, 2001a, 2001b) their role in the grammar is to drive syntactic operations, particularly movement.

Chomsky (2001a) furthermore proposes that the formal difference between the interpretable features and their uninterpretable counterparts is that the latter enter the derivation unspecified, being assigned values as part of the process of derivation by virtue of entering into the relation Agree with an interpretable counterpart. See Zwart 1997: 189 for an early version of the same idea. This theory gives formal expression to the intuition that agreement is directional. In for example (7), the auxiliary verb and the adjective agree with the subject NP, not vice versa. Once the uninterpretable features are assigned values, they are removed from the syntactic derivation being handed over to morphology/phonology, the derivation of PF.
3. Two hypotheses

Within this theory of agreement it is obviously not possible for an inherently unspecified pronoun to be specified by the $\phi$–features of I, as those features are themselves inherently unspecified. The following are two alternative hypotheses consistent with the feature theory sketched above:

_Hypothesis A:_

There is no _pro_ at all in null-subject constructions. Instead Agr, the set of $\phi$–features of I, is itself interpretable; Agr is a referential, definite pronoun, albeit a pronoun phonologically expressed as an affix. As such, Agr is also assigned a subject theta-role, possibly by virtue of heading a chain where the foot of the chain is in vP, receiving the relevant theta-role.

Versions of this hypothesis are articulated in Alexiadou & Anagnostopoulou 1998, Manzini & Roussou 1999, Manzini & Savoia 2002, and Platzack 2003, 2004. If Agr is interpretable, it could, on that account, specify the features of _pro_. But if Agr is interpretable, there is no need for _pro_. The role of (subject) _pro_ in Chomsky 1982, Rizzi 1986 and related work is to carry the subject theta-role, possibly nominative Case, and satisfy the EPP. But if Agr is interpretable, hence referential, then Agr may itself carry the subject theta-role. This means that there could at most be an expletive _pro_ in specIP. If Agr absorbs nominative Case as well, as seems most plausible if it is referential and heads a chain, then it would be a Caseless expletive _pro_. Expletive _pro_ is a dubious category, particularly in a minimalist framework, as it has no interface properties at all, neither at LF nor PF. But even granting the theoretical possibility, the only condition that could conceivably require an expletive _pro_ in specIP would be the EPP. Alexiadou & Anagnostopoulou 1998 and Manzini & Savoia 2002
exclude this possibility by stipulating that the EPP is (effectively) satisfied by Agr in null-subject languages. We may, however, for the sake of argument, retain as a theoretical possibility that a covert expletive pro can satisfy the EPP in a null-subject language.

**Hypothesis B:**

The null subject is specified for interpretable $\phi$-features, values the uninterpretable features of Agr, and moves to specIP, just like any other subject. This implies that the nullness is a phonological matter: The null subject is a pronoun which is not pronounced.

In the following I will, for the sake of argument, ignore the logical possibility that some languages do not have an EPP-requirement. With this proviso, the following is an empirical difference between hypotheses A and B: According to A, in finite null-subject constructions the subject position specIP is either not projected or filled with expletive pro, the former if Agr on the finite verb can check (satisfy) the EPP, the latter if it cannot. According to B, the position is occupied by a pronoun checking the EPP, and is therefore not available for another category.

Now assume a language which allows null subjects, but has an overt expletive. Hypothesis A is consistent with the following three alternatives: The overt expletive is (a) excluded, (b) allowed, or (c) compulsory in finite null-subject constructions. Alternative (a) would hold if interpretable Agr necessarily checks the EPP. Alternative (b) would hold if interpretable Agr may, but need not check the EPP. Alternative (c), finally, would hold if Agr, interpretable or not, cannot check the EPP. Hypothesis B, on the other hand, categorically excludes the overt expletive from occurring in null-subject constructions.
Null-subject languages are generally assumed not to have an overt expletive pronoun, especially not a ‘pure’, nominal expletive such as English *there.* There is at least one null-subject language which has an overt pure, nominal expletive, though, namely Finnish, as discussed by Holmberg & Nikanne 2002.

I will initially put aside the alternative that interpretable Agr itself checks the EPP and thereby excludes the overt expletive in null-subject constructions, and come back to this version of Hypothesis A in section 6. With this proviso we can use Finnish to decide between the competing hypotheses A and B: Hypothesis A allows the overt expletive to occur in null-subject constructions, Hypothesis B excludes it. As will be demonstrated below, Hypothesis B makes the right prediction for Finnish.

4. Null subjects and agreement in Finnish

Finnish is a partial null-subject language in that 1st and 2nd person pronouns are optionally null, in any environment.

\[(\text{Minä) puhun englantia.}) \quad (\text{Me) puhumme englantia.})\]
\[
\text{I speak-1SG English} \quad \text{we speak-1PL English}
\]
\[(\text{Sinä) puhut englantia.}) \quad (\text{Te) puhutte englantia.})\]
\[
\text{you speak-2SG English} \quad \text{you-PL speak English}
\]
\[
*(\text{Hän) puhuu englantia.}) \quad *(\text{He) puhuvat englantia}
\]
\[
\text{he/she speak-3SG English} \quad \text{they speak-3PL English}
\]

A 3rd person definite subject pronoun can be null when it is bound by a higher argument, under conditions which are rather poorly understood.
a. Pekka väittää [että hän/,Ø/ j/ puhuu englantia hyvin].

Pekka claims that he/Ø speaks English well

b. Anu sanoi Jarille, että hän/,Ø/ j/ ottaa kitaran mukaan.

Anu said Jari-ALL that he takes guitar along

‘Anu told Jari to bring along his guitar.’

c. Se oli Tarjalle pettymys [ettei hän/,Ø/ j/ saanut lukea latinaa koulussa].

it was Tarja-ALL disappointment that-not she/Ø could study Latin school-INE

‘It was a disappointment for Tarja that she couldn’t study Latin at school.’

d. Poikien mielestä oli noloa kun hän/,Ø/ j/ jäivät kilpailussa viimeiseksi.

boys’ opinion-ABL was embarrassing when they came race-INE last

‘The boys found it embarrassing when they came last in the race.’

e. Jokaisen pojan mielestä on noloa kun hän/,Ø/ j/ jää kilpailussa viimeiseksi.

every boy’s opinion-ABL is embarrassing when he comes race-INE last

‘Every boy finds it embarrassing when he comes last in a race.’

f. Se oli Tarjaniäidille pettymys [ettei hän/,Ø/ j/ saanut lukea latinaa koulussa].

it was Tarja’s mother-ALL disappointment that-not she/Ø could study Latin school-INE

‘It was a disappointment to Tarja’s mother that she could not study Latin at school.’
Vainikka & Levy 1999 assert that the embedded null 3rd person must be coreferent with an argument in the next clause up. This is echoed in Gutman 2004, who discusses conditions on the null subject-antecedent relation on the basis of examples with a subject, object, and indirect object antecedent. Examples such as (9c,d) show that the structural conditions on the relation are quite lax: In (9c) the antecedent is an adjunct, while in (9d) it is embedded in an NP (which is embedded in a PP, if, following Nikanne (1993), we take the Finnish locative cases, including ablative, to be assigned by a covert adposition), hence does not c-command the null subject. (9e) indicates that the relation between a quantified argument and a null variable in the subject position of an embedded finite clause is subject to similar lax structural conditions. (9f,g), finally, show that the relation is still subject to stricter conditions than that between an overt pronoun and its antecedent: The null subject cannot support coreference, in the manner of a Principle B pronoun. Comparison of (9d,e) suggests that the relation between the null subject and the higher referential DP in the examples in (9) is a special case of variable-binding, where the variable may be null. I will henceforth refer to the relation as binding, with the understanding that the structural conditions on the relation are not as strict as in Chomsky’s (1991, 1982) Binding Theory.4

Generic pronouns can, and must, be null.

(10) Täällä ei saa polttaa.

here not may smoke
‘One can’t smoke here.’

Quasi-referential subjects in construction with extraposed clauses can also be null, and, with certain exceptions, must be null in construction with weather predicates; see Holmberg & Nikanne 2002.

(11) a. (Se) oli hauskaa että tulit käymään.

    it was nice that came-2PL visiting

    ‘It was nice that you came to visit.’

b. Sataa vettä.

    rains water

    ‘It’s raining.’

As shown by (8), Finnish has rich agreement morphology. 3rd person agreement is less rich than 1\textsuperscript{st} and 2\textsuperscript{nd} person agreement in that 3SG is null in the past tense and in the conditional mood where tense is neutralized; see Holmberg & Nikanne 1993, Holmberg & al. 1993. In many varieties of colloquial Finnish there is no distinction between 3SG and 3PL. In these varieties, 3PL is also null in the past tense and the conditional mood. In the present tense indicative 3SG is, however, phonologically visible in the form of vowel lengthening, and 3PL has a suffix –vät/–väät (subject to vowel harmony) in all tenses and moods, except in those colloquial varieties where it is = 3SG.

The use of 1\textsuperscript{st} and 2\textsuperscript{nd} person null subjects is largely restricted to formal varieties of Finnish, including standard written Finnish. It is nonetheless clear that these null subjects are part and parcel of Finnish ‘core grammar’, since Finnish speakers have largely uniform intuitions about null-subject constructions. See Holmberg & Nikanne 2002 for discussion.
The use of 3rd person null subjects in cases like (9) is not restricted by such stylistic considerations.

5. The Finnish expletive sitä

Finnish has an expletive pronoun which is obligatory in certain contexts. In general, Finnish does not tolerate verb-initial declarative sentences, hence (12a) is ungrammatical. You either have to move and remerge a referential category (an argument or a referential adverbial) with IP, or merge the expletive sitä.

(12)

a. *Sattui minulle onnettomuus.
   happened to-me accident

b. Minulle sattui onnettomuus.
   to-me happened accident

c. Sitä sattui minulle onnettomuus.
   EXP happened to-me accident
   'I had an accident.'

The same pattern is illustrated in (13) and (14).

(13)

   went now wrong

b. Nyt meni hullusti.
now went wrong

c. **Sitä** meni nyt hullusi.

EXP went now wrong

'Now things went wrong.'

(14) a. *Viihtyy saunassa.

feels-good in-sauna

b. **Saunassa** viihtyy.

in-sauna feels-good

c. **Sitä** viihtyy saunassa.

EXP feels-good in-sauna

'One feels good in the sauna.'

The expletive sitä is the partitive form of the pronoun se ‘it’. As shown in Holmberg & Nikanne 2002, it is a *there*-type, pure expletive, merged in the position where the subject is found in unmarked sentences. As also shown by Holmberg & Nikanne, this is not a subject position *per se*, as it can be filled by other referential categories. Holmberg & Nikanne characterize it as a topic position, which is, however, shown to be not quite correct in Holmberg (to appear). I will continue to refer to it as specIP. It is not a specCP position, as shown, for instance, by the fact that it can invert with the finite verb in questions, where the finite auxiliary or verb moves to C, and is affixed with a question particle:
(15) Menikö sitä taas hullusti?
   went-Q EXP again wrong
   ’Did things go wrong again?’

There is not necessarily any interpretive difference between the (b) and the (c)-sentences in (12, 13, 14). Fronting the argument or temporal or locative adjunct is the solution to a formal condition, the Finnish version of the EPP, as is merge of the expletive. The fronted argument or adjunct can have more specific information-structural implications, but need not have any. In fact, in written Finnish the use of the expletive is proscribed, leaving fronting as the only acceptable means to satisfy the EPP.

A couple of caveats are in order: First, a verb-initial sentence is acceptable with ‘verb-focus’, or rather polarity focus, as in (16), where the verb tends to have focus stress and possibly a suffixed focus particle.

(16) SATTUI(-pas) minulle onnettomuus.
    happened (FOC) me-ALL accident
    ‘I did have an accident.’

In this case the finite auxiliary or verb, or more precisely, the functional head which incorporates the finite auxiliary or verb and encodes polarity is moved to C: see Holmberg 2001.

Second, verb-initial impersonal sentences are allowed if the sentence contains no category that can move to specIP to check the EPP. A category can check the EPP (a) if it is a subject or (b) if it is referential, in the sense that DPs and certain adverbials (locative, temporal, instrumental, but not for instance manner or reason) are referential; see Holmberg
Adapting Holmberg & Nikanne (2002), let us say a non-subject must be a ‘potential topic’ to check the EPP. For instance (16) is therefore acceptable. Compare (13a) and (17):

(17) Meni hullusti.

\begin{quote}
went wrong
\end{quote}

‘Things went wrong’

The manner adverb *hullusti* is not a potential topic, while the time adverb *nyt* ‘now’ in (13) is, hence the difference: The time adverbial must move to specIP in the absence of an expletive. See Holmberg & Nikanne 2002 for discussion of verb-initial clauses and the EPP in Finnish. The following is thus a viable formulation of the Finnish version of the EPP:\(^5\)

(18) **EPP in Finnish**: If the sentence contains one or more categories which can check the EPP, then one of them must remerge with IP, or an expletive be merged with IP.

It is important to note that the expletive is not restricted to impersonal and generic sentences, but also occurs in construction with, for instance, 1\(^{st}\) and 2\(^{nd}\) person finite verbs and ditto subjects, if the latter are not remerged with IP, as in (19):\(^6\)

(19)a. Sitä olen minäkin käynyt Pariisissa.

\begin{quote}
EXP have-1SG I-too visited Paris-INE
\end{quote}

‘I have been to Paris, too (actually).’
b. Minä sitä olen käynyt Pariisissa.

I EXP have-1PL visited Paris-INE

‘I’ve been to Paris (would you believe it).’/ ‘I’m the one who has been to Paris.’

In (19a) the subject pronoun is in a sentence-medial focus position with the expletive in specIP checking the EPP. In (19b) the subject pronoun has moved to the sentence-initial focus position (a specCP position; see Vilkuna 1995, Holmberg & Nikanne 2002), in which case, again, the expletive can check the EPP. Note also that the absence of a referential category which can check the EPP, the expletive is optional; compare (17) and (20).

(20) Sitä meni hullusti.

EXP went awry

‘Things went wrong.’

6. Testing hypotheses A and B

The stage is now set for testing the two hypotheses presented in section 3. According to Hypothesis A, Finnish 1st and 2nd person Agr is made up of interpretable features, and so is essentially an affixed definite pronoun. An overt 1st or 2nd person subject pronoun is therefore not required, and if included, it is (presumably) in a higher, A-bar-type position. The prediction is, then, that specIP, the position immediately preceding the finite verb or auxiliary in a declarative sentence, or immediately following it in a yes/no question, could, or even should be filled with an expletive pronoun. The prediction is false, as first observed in Hakulinen 1975; see also Holmberg & Nikanne 2002.
(21)a. *Sitä puhun englantia.
   EXP speak-1SG English

   b. Oletteko (*sitä) käyneet Pariisissa?.
      have-2PL-Q EXP visited Paris
      'Have you been to Paris?'

This is predicted by Hypothesis B, according to which a null pronoun checks the EPP. (21a,b), with the expletive, are ill formed for the same reason that (22a,b) are: The subject pronoun checks the EPP, thus leaving no function for the expletive to fulfill.

(22)a. *Sitä minä puhun englantia.
   EXP I speak-1SG English

   b. Oletteko te (*sitä) käyneet Pariisissa?
      have-2PL you EXP visited Paris

(22a,b) compared with (19a,b) show that a subject pronoun can’t be combined with an expletive in pre-verbal position in either order (pronoun-expletive or expletive-pronoun) except if the pronoun is focused, in which case it may occupy a position other than specIP, leaving it up to the expletive to satisfy the EPP.  

The same point, that null 1st and 2nd person subjects in Finnish are syntactically represented in specIP (in their terms, specAgrSP), is made by Vainikka & Levy 1999, on the basis of pairs such as the following:
(23a) shows that a locative phrase can satisfy the EPP (in our terms), when the subject is postverbal. (23b) shows that this is not possible in construction with a null 1PL subject. The examples are selected so as to exclude fronting to specCP as an analysis of the fronted locative. The conclusion is that the null subject in (23b) occupies the EPP-position, that is specIP in our terms.

7. A counterproposal

There is another interpretation of the facts, particularly (21a,b), which is compatible with Hypothesis A: The subject agreement category Agr in a null-subject language, in Finnish particularly 1st and 2nd person Agr, is a referential, definite pronoun which is incorporated in I but is nevertheless capable of checking the EPP. Therefore merge of an expletive is precluded in (21), since the only function of an expletive is to check the EPP. This would accord with Alexiadou & Anagnostopoulou’s (1997) and Platzack’s (2003, 2004) theories of null subjects.

There are good reasons to reject this account of (21a,b). To begin with, it would wrongly exclude sentences such as (19a,b), repeated here:

(19a) Sitä olen minä-kin käynyt Pariisissa.
EXP have-1SG I-too visited Paris-INE

‘I have been to Paris, too (actually).’

b. Minä sitä olen käynyt Pariisissa.

I EXP have-1SG visited Paris-INE

‘I’ve been to Paris (would you believe it).’ / ‘I’m the one who has been to Paris.’

If 1SG Agr checks the EPP, then what is the expletive doing in (19a,b)? On the other hand, if the EPP needs a category in specIP, then the role of the expletive in (19a,b) is to check the EPP when the thematic subject is not checking it because it is in a focus position elsewhere.

The expletive is optional in (19a,b), in the sense that the sentences are well formed without it, even with preserved focus on the subject.

(24)a. Olen minäkin käynyt Pariisissa.

be-1SG I-too visited Paris-INE

‘I have been to Paris, too.’

b. MINÄ olen käynyt Pariisissa.

I be-1SG visited Paris-INE

‘I have been to Paris.’ / ‘I’m the one who has been to Paris.’

This suggests that Agr checks the EPP optionally. But in that case (21a,b) are wrongly predicted to be well formed, that is when the option not to have Agr check the EPP is taken. Instead, I claim, the difference between (24a,b) and (19a,b) is that the subject pronoun itself checks the EPP in (24a,b). This is straightforward in (24b): In the absence of an expletive, the
subject merges with IP checking the EPP on its way to specCP. (24a), I propose, is derived by merging the subject with IP, checking the EPP, and moving the auxiliary, incorporated in finite I, to C; see (16) above.  

Furthermore, if the only reason for subject movement to specIP is the EPP (following Chomsky 2000, 2001), and if Agr checks the EPP-feature of I, then a subject preceding finite I must be in an A'-position, the movement triggered by a feature in the C-domain, typically a feature with information-structural import. This is argued to be a correct prediction by Alexiadou & Anagnostopoulou (1997) and Platzack (2003, 2004) for the null-subject languages they discuss. The prediction is incorrect for Finnish, even aside from the fact that an expletive may precede finite I in this language.

As discussed by Vilkuna (1989, 1995), there are strictly two positions in the left periphery preceding the finite verb or auxiliary, in Finnish: A contrastive position which is also the landing site for a fronted wh-phrase, and a position which in the unmarked case is occupied by the subject, but may be occupied by any category capable of a topic interpretation. Vilkuna (1989) terms the positions ‘K’, suggesting ‘contrast’, and ‘T’, suggesting ‘topic’. For example, (25a) (adapted from Vilkuna 1995) can only be interpreted with Annalle ‘to Anna’ as contrastive focus. (25b) can only be interpreted with the fronted object kukkia as contrastive focus and Annalle as topic (the postverbal subject in that case being information focus or ‘Main News’ in Vilkuna’s 1995 terms), while (25c) is ill formed, having one argument too many preceding the finite verb; see Vilkuna 1995.


Anna-ALL Mikko gave flowers

’It was to Anna that Mikko gave flowers.’
b. Kukkia Annalle antoi Mikko
   flowers Anna-ALL gave Mikko
   ’Flowers, Anna received from Mikko.’

c. *Kukkia Annalle Mikko antoi. 10
   flowers Anna-ALL Mikko gave

It does not matter whether the arguments are pronouns or lexical NPs; the information-structural interpretation is the same. Vilkuna (1989, 1995) and also Holmberg & Nikanne (2002) assume that a preverbal argument which is not contrastive is always topic, in terms of information structure. This cannot be right, since the subject can be for instance an indeterminate pronoun, not a possible topic, as in (26) (the negation is a finite auxiliary in Finnish). It still holds true that only one constituent can precede the subject, and it must be contrastive (or a whP).

(26) a. Annalle kukaan ei antaisi kukkan.
    Anna-ALL anybody not-3SG give-CON flowers
    ‘Nobody would give flowers to ANNA.’

b. *Annalle kukkan kukaan ei antaisi.

The generalization can be stated as follows:

(27) The finite verb or auxiliary (including the negation) can be preceded by at most two sentence constituents: The one closest to the finite verb or auxiliary checks the EPP, the other is contrastive (or a whP).
If 1\textsuperscript{st} and 2\textsuperscript{nd} person Agr checks the EPP in Finnish, the prediction is that an overt pronoun preceding Agr should have contrastive interpretation.

(28) Minä olen käynyt Pariisissa.

I be-1SG visited Paris-INE

‘I have been to Paris.’

The prediction is false. The initial pronoun can be, but certainly need not be in the specCP contrast position. Thus it can be preceded by a contrastive category, and it can be the subject of a conditional clause.

(29) a Pariisissa minä olen käynyt (mutten Roomassa).\textsuperscript{11}

Paris-INE I be-1SG visited but-not Rome-INE

‘I’ve been to PARIS (but not Rome).’

b. Jos minä olisin käynyt Pariisissa, …

if I be-CON-1SG visited Paris-INE

‘If I had been to Paris, …’

We may conclude that a finite 1\textsuperscript{st} or 2\textsuperscript{nd} person verb or auxiliary in Finnish does not check the EPP. Instead, it takes a category merged with IP to check the EPP, either by movement or by merging an expletive. In the Finnish null-subject construction, therefore, there is a null subject pronoun in specIP. Following the Chomskyan approach to agreement, the null pronoun has interpretable φ–features, and assigns values to the inherently unvalued
features of Agr. In other words, the null subject pronoun identifies Agr (= the finite verb or auxiliary agrees with the null pronoun), not vice versa.

In the rest of the paper I will argue for a typology of null subjects. Section 5 will present the basic properties of the Finnish generic null subject. Section 6 will deal with the Finnish null bound subject. It will be shown that, while both are syntactically projected, they occupy different structural positions. I will then deal with null subjects in consistent null-subject languages such as Spanish, and finally null subjects in Agr-less languages such as Japanese.

8. The null generic pronoun in partial null-subject languages

As mentioned, Finnish has a null generic pronoun.

(30)  
  a. Tässä istuu mukavasti.
       here sit-3SG comfortably
       ‘One can sit comfortably here.’

  b. Opettajana odottaisi vähän kunnioitusta.  (Laitinen 1995)
       teacher-ESS expect-CON-3SG some respect
       ‘As a teacher one would expect a bit of respect.’

Unlike definite null pronouns, the generic null pronoun does not count for the EPP. Consequently fronting of the locative in (30a) and the essive adjunct in (30b) is compulsory, required to check the EPP. Alternatively the expletive sitä can be merged to check the EPP. The corresponding sentences in (30) and (31) are synonymous.
This suggests that the generic subject might not be syntactically projected. However, scholars who have investigated the Finnish generic subject construction (GSC), including Hakulinen & Karttunen (1973), Vainikka (1989), Laitinen (1995), Vainikka & Levy (1999), Holmberg (to appear), agree that it contains a syntactically represented subject. One often mentioned piece of evidence is that it can be the antecedent of an anaphor.

Important in this connection is the insight, first discussed by Vainikka (1989), that the 3SG value of the finite verb/auxiliary in the examples above is not default 3SG, but is assigned by a nominative subject. The evidence is that the object in, for example, (33) is assigned accusative case, marked by –n in the singular.
You can buy a car here.

Simplifying somewhat, the rule in Finnish is that the object of a transitive verb gets assigned accusative if and only if the first finite sentence dominating the object has a subject with which the finite verb or auxiliary agrees, which is to say the subject must have nominative case; see Timberlake 1972, Maling 1993, Reime 1993, Nelson 1998, Kiparsky 2001.12 In (34) the subject is assigned genitive, a lexical case assigned by necessive verbs to their subject (see Laitinen & Vilkuna 1993). Consequently the verb does not agree with the subject, but has default 3SG form, and consequently the object of the main verb has nominative case.13

(34) Meidän täytyy ostaa *auton/ auto.

we-GEN must-3SG buy car-ACC/car-NOM

‘We must buy a car.’

Now consider (35a,b):

(35)  a. Täällä voi ostaa auton/*auto.

here can-3SG buy car-ACC/car-NOM

‘One can buy a car here.’

b. Nyt täytyy ostaa auto/*auton.

now must-3SG buy car-NOM/car-ACC

‘One must buy a car now.’
In (35a) the object has accusative case, which presupposes that the sentence contains a nominative subject assigning values to Agr. This will be the case if the sentence contains a null generic subject pronoun assigned nominative by T. In (35b) the object has nominative case, which presupposes that the sentence does not contain a nominative subject assigning values to Agr. This will be the case if the sentence contains a null generic pronoun to which the necessive verb assigns genitive case. The 3SG form of the verb töytyy is the default finite indicative form.

Alternative accounts may be considered. One, compatible with Hypothesis A, is that 3SG Agr is an interpretable incorporated pronoun, which in Finnish is not a referential but a generic pronoun (or an impersonal pronoun interpretable as generic). If so, there is no null generic pronoun. But this theory will have a problem accounting for the facts in (35a,b). Within that theory, it would appear, there is no interpretable 3SG Agr in I in a case like (34), as there is no agreement (the verb just has its default finite form). But in that case there is no subject at all in (35b), which ought to cause a theta-criterion violation, and is also contradicted by the observation that the object may be an anaphor.

(36) Nyt töytyy pestä auto-nsa.

now must wash car -POSS.RFL

‘One must wash one’s car now.’

See Holmberg (to appear) for additional arguments against this alternative theory.
9. The null logophoric 3rd person pronoun

Why does the null generic subject not check the EPP? The fact that it is an impersonal, indefinite category, similar to, although not synonymous with the quantifier anyone, is not a sufficient explanation, since, as mentioned in section 5, any subject, whatever its semantic properties, can check the EPP.

In this connection, consider (37), first observed by Hakulinen (1977), also discussed in Vainikka & Levy 1999: 648.

(37) a. Oppilas tietää ettei pysty ratkaisemaan tehtävää.
student knows that-not can solve assignment
‘The student knows that he can’t solve the assignment.’

b. Oppilas tietää ettei tehtävää pysty ratkaisemaan.
student knows that-not assignment can solve
‘The student knows that the assignment can’t be solved.’

Recall that, although Finnish does not have a null definite 3rd person subject in main clauses, it does in embedded (finite) clauses if the null subject is bound by an argument in the next clause up (see section 4). I will refer to this as a null logophoric subject pronoun. It appears to conform to Sell’s (1987) characterization of a logophoric pronoun, being typically found in embedded clauses reporting the attitude or mental state of the subject or other argument of the superordinate clause, binding the embedded null subject. The logophoric reading of the null subject requires, however, that no argumental category such as an object or a locative
adverbial precedes the embedded finite verb/auxiliary. If that is the case, the only possible reading is that the subject of the embedded clause is generic.\textsuperscript{15}

I propose that the logophoric null pronoun and the generic null pronoun are the same category. I propose, furthermore that it is a $\phi$P. Since it values Agr, as shown in the preceding section, it must have inherently valued (hence interpretable) $\phi$-features. Crucially, though, it lacks the substructure required for a definite, referential category. In the spirit of Longobardi 1994 I take the lacking property to be the head D, in the absence of which the pronoun cannot refer to an individual or group, neither independently/deictically nor under coreference with an independently referring DP. However, it can be a variable bound by a QP, or be ‘logophorically linked’ to a DP in a higher clause. As a last resort, it can be interpreted as generic.

The label ‘$\phi$P’ for a subcategory of pronouns comes from Dechaine & Wiltschko 2003 (henceforth D&W), who propose a typology where pronouns are either DPs, $\phi$Ps, or NPs.\textsuperscript{16} Their $\phi$P is similar to the one assumed here, but not identical. A $\phi$P in D&W can both support coreference and function as a bound variable. This makes, for example, the English 3\textsuperscript{rd} person pronouns, and also the Finnish 3\textsuperscript{rd} person overt pronouns, $\phi$Ps.\textsuperscript{17} The Finnish 3\textsuperscript{rd} person null pronoun does not, however, qualify as a $\phi$Ps in D&W as it cannot support ‘ordinary’ coreference (that is, except as a logophor). ‘Our’ $\phi$P has less structure than $\phi$P in D&W. As will be discussed below, the missing structure (or feature) is provided by I in some languages, but not in Finnish.\textsuperscript{18}

The null $\phi$P in Finnish is accessible for binding by a higher DP if and only if it moves to specIP. If it remains in specvP it is inaccessible, and the generic interpretation is the only option. See Vainikka 1989: 234f. and Holmberg, to appear, for arguments that it does remain in specvP. The reason why the $\phi$P is accessible when moved out of vP is clear in principle: It moves to position closer to the root of the sentence, and therefore closer to the antecedent.
The precise implementation of this idea is not so straightforward, though. Ideally it ought to follow from phase theory (Chomsky 2000, 2001a,b): The $\phi$P has to move to the edge of a phase to be accessible for a DP in the next higher phase. However, no current version of phase theory takes IP (TP) to be a phase. I will leave the precise formal account for future research, confident that the generalization that the logophoric reading is possible only when the null $\phi$P subject occurs higher up in the structure can receive a satisfactory explanation in terms of a theory of locality, perhaps a version of phase theory.

Why $\phi$P cannot be interpreted as generic except if it stays in specvP is rather less obvious. The following is a possibility: The generic interpretation is the result of an indefinite expression (the $\phi$P) being bound by an abstract generic operator (see Krifka & al. 1995). If the generic operator is located no higher than I, then movement of $\phi$P to specIP or higher will move it out of the c-command domain of the operator, ruling out a generic interpretation, leaving the logophoric interpretation as the only option. A problem for this hypothesis is that non-pronominal generic subjects are not as a rule confined to specvP. Instead, as discussed by Diesing (1992), generic arguments, even when they are bare $\phi$Ps (bare plurals in the case of Germanic) have a syntactic distribution similar to definite arguments, typically occurring in higher positions than indefinite arguments. In Finnish, too, there is no indication that a non-pronominal generic subject would, in general, have to be lower than I. For example, in (38) the generic subject, on the face of it a bare noun, is in specIP, the usual subject position preceding the negation.

(38) Tiikeri ei kiive puihin.
    tiger not climb trees-INE
    The tiger doesn’t climb trees.
It seems to be the case, then, that the null bare $\phi P$ has other properties than overt generic arguments.

The hypothesis that the generic null subject and the logophoric null subject in Finnish are instances of the same category is supported by the observation that we find the same array of properties in a number of other languages, unrelated to Finnish. Consider first Brazilian Portuguese:

(39) a. Ele/*Ø ganhou na loto. (Brazilian Portuguese)
   he won on the lottery

   b. Pedro, disse que ele/*i/j ganhou na loto.
   Pedro said that he won on the lottery

   c. Aqui não pode nadar.
   here not can swim

   ‘One can’t swim here.’

(39a) exemplifies the fact that Brazilian Portuguese does not allow a 3rd person null subject in a main clause. (39b) shows that it does allow it in a finite embedded clause, when the null subject is logophorically dependent on the subject in the next higher clause. (39c), finally, shows that Brazilian Portuguese has a generic null subject.

Next, consider Marathi:19

(40) a. to/*Ø lotteri jink-l-a
   he /*Ø lottery win-Past-3M.SG

   'He won the lottery'
b. Ram\textsubscript{i} mhana-\textsubscript{l}-a \text{ki to}\textsubscript{\textit{\textalpha}}/\emptyset_/\textsubscript{\texti}\text{lotteri jink-\textsubscript{l}-a}  
R say-Past-\textit{\textbf{3M}s that he /\emptyset_/ lottery win-Past-\textit{\textbf{3M}}.\textit{SG}  
‘Ram said that he won the lottery’

c. Hya khurchi- war aaramani bushushakto.  
this chair -on comfort-with sit-PRES.\textit{3SG}  
‘One can sit comfortably in this chair.’

Again (40a) shows that Marathi does not allow a 3\textsuperscript{rd} person null subject in a main clause. (40b) shows that it does allow a null subject bound by the subject in the next clause up. (40c), finally, shows that Marathi has a generic null subject.

Finally, (41) exemplifies the same array of properties in Hebrew::

(41) a. Hu/\emptyset_/’axa\textsubscript{l} ’et ha-tapu’\textsubscript{ax}. (adapted from Borer 1986)  
he ate-\textit{3SG} ACC the-apple  
‘He ate the apple.’

b. Talila\textsubscript{i} ’amra le-Itamar\textsubscript{j} she hi\textsubscript{i}/hu\textsubscript{j}/\emptyset_/\textsubscript{\texti}/\textk tavo. (adapted from Borer 1986)  
Talila said to-Itamar that she /he /\emptyset_/ will-come-F-\textit{SG}  

c. Yxolim la-\textsubscript{\textsigma}e\textsubscript{\textsigma}v et be-noxiout ba-kise ha-ze.  
can-\textit{3PL to-sit in-comfort in-the-chair the-this}  
‘One can sit comfortably in this chair.’
There are certain differences among these languages, regarding null subjects in main clauses; for instance, Finnish and Hebrew allow 1st and 2nd person null subjects, Marathi only 2nd person null subjects, while Brazilian Portuguese doesn’t allow any. In Hebrew the agreement in the generic constructions is 3PL, in the other languages 3SG. There are also differences regarding the required structural relation between the embedded null subject and the antecedent. Nevertheless, the similarities among these partial null-subject languages are more striking than the differences, when they are compared with consistent null subject languages. As discussed in Holmberg (to appear), languages which allow a definite 3rd person null subject in main as well as embedded clauses do not have a generic 3rd person null subject corresponding to English ‘one’.

Instead, to express the meaning of generic ‘one’ they resort to some form of overt morphology, such as (cognates of) the reflexive se in Romance and (most varieties of) Slavic, as in the following European Portuguese example.

(42) Aqui não se pode nadar. (European Portuguese)
    here not SE can swim
    ‘One can’t swim here.’

Alternatively they resort to generic ‘you’ (which may be null, but with 2SG agreement on the finite verb), overt quantifiers such as ‘anyone’, or a variety of other strategies to avoid the use of a null, generic 3rd person pronoun. See Holmberg (to appear) for examples and discussion.

The clustering of the three properties exemplified in (39-41) across languages is consistent with the hypothesis that the null generic pronoun and the null logophoric pronoun are the same category; a category which is furthermore incompatible with a definite 3rd person null subject. A formal explanation of the clustering of the properties is proposed in the next section.
10. The null subject in consistent null subject languages

Why do languages have to choose between a definite 3\textsuperscript{rd} person null subject and a generic 3\textsuperscript{rd} person null subject?

I propose that consistent null subject languages such as Spanish have a D-feature in I, which is lacking in Finnish and the other partial null subject languages listed in the previous section. The idea that finiteness involves a sentential D-feature recurs in various forms in recent literature; see especially Chomsky 1995: 282 and Alexiadou & Anagnostopoulou 1998. I propose that the D-feature is parametrized in the following way: Presence of a D-feature in I means that a null \( \phi \)P which enters into an Agree relation with I can be interpreted as definite, referring to an individual or a group. Furthermore, following Holmberg (to appear) I assume it means that a null subject cannot be interpreted as generic. Absence of D in I, on the other hand, means that a null \( \phi \)P subject must be either bound by a higher DP, or else be interpreted as generic.

In this connection, consider Cardinaletti & Starke’s (1999) (C&S) theory of pronouns: According to C&S UG provides for three types of pronouns: strong, weak and clitic. Strong pronouns are full CPs, the pronominal equivalent of sentential CPs, which means that they have the structure required for independent reference. The property which C has which makes this possible is the feature K, short for functional case, which is the syntactic counterpart of a referential index (“index is the interpretation of K”; C&S: p. 190). Weak and clitic pronouns are both deficient in that they lack C, hence K. This has the consequence that they have to move from their original theta-position to the spec of a case-assigning head, that is AgrO for objects, AgrS for subjects. Clitic pronouns have even less functional structure than weak pronouns, and therefore have to undergo further movement, with adjunction to a head.\textsuperscript{22} Their
lack of K then explains why weak and clitic pronouns, as opposed to strong pronouns, appear to always undergo movement. Entering a spec-head relation with Agr compensates for the lack of K in weak and clitic pronouns. Cardinaletti & Starkes’s theory is not couched in terms of (un-)interpretable features, but the implicit assumption is that Agr is interpretable, providing the deficient pronouns with the K-feature required for (co-)referential interpretation.

Furthermore, and importantly for the present theory, Cardinaletti & Starke argue convincingly that null arguments are weak pronouns.

I will modify their theory slightly: Focusing on subject pronouns, Agr, the set of ϕ-features realized as subject agreement, is not the category which compensates for the lack of K in deficient pronouns, assigning them referential capacity. It cannot be, if Agr is uninterpretable and assigned its values by a nominal argument, be it a (pronominal) CP or a deficient pronoun. Instead the crucial feature is a feature independent of Agr which some languages but not others have, co-existing with Agr as a component of I. Remaining agnostic regarding the precise functional structure of nominal arguments, and the relation of case to referentiality, I will continue to use the more traditional label ‘D’ for the feature which distinguishes arguments inherently capable of reference from referentially deficient arguments.

Consistent null subject languages such as Spanish, European Portuguese, Greek, Turkish, etc. have a D-feature in I. A definite null subject is a ϕP, a deficient pronoun which receives capacity to refer to an individual or a group from I containing D.

We may assume that the relation between ϕP and D is an Agree-relation: ϕP has a feature [uD] (unvalued D) which is valued either by merging D with ϕP, which yields [DP D ϕP], or by merging D as a component of I in a local c-commanding relation to ϕP merged with vP (where ϕP usually ends up remerged with IP). If ϕP and its [uD] feature is not locally c-
commanded by D, it can still be licit if it is bound by a DP (subject to parametric variation), or interpreted as generic. Valueing [uD] precludes a generic interpretation of the null subject. Therefore consistent null subject languages have to resort to a variety of ‘overt strategies’ to express the meaning of a generic subject.pronoun; see Holmberg (to appear).23 24

Where does this leave 1st and 2nd person null subjects in Finnish? As they assign person and number feature values to Agr, they must encode person and number. As they are interpreted as definite, and the definiteness cannot be ascribed to a feature in I, as Finnish, by hypothesis, lacks such a feature, they must be DPs. We are led to conclude that the 1st and 2nd person null subjects are fully specified DP pronouns which are deleted, presumably by essentially the same process that applies in other well known cases of ellipsis, such as VP-ellipsis, NP-ellipsis, etc. Recoverability is ensured by the agreement marking on I..

If this is correct there are two types of definite null subjects: One is an inherently null deficient pronoun which has to enter an Agree-relation with I containing D to be interpreted as a definite argument (lacking descriptive content, it will still usually be dependent on an antecedent to have its reference fixed). It can also be interpreted as a bound variable pronoun or a logophoric pronoun. In the absence of D in I, it can still be interpreted as a bound or logophoric pronoun, or, in the absence of a binder, as a generic pronoun. The other type is a fully specified DP which is deleted.

In Spanish (Greek, Turkish, etc.) all null subjects are of the former, deficient, type, while in Finnish only 3rd person null subjects are; 1st and 2nd person null subjects are fully specified DPs.

11. Non-null-subject languages

We can distinguish three types of languages:
(43)  
A. Consistent null-subject languages;  
B. Partial null-subject languages;  
C. Non-null-subject languages.

The difference between languages A and B is that languages A have a D-feature in I, absent in languages B. What about languages C? Since, like languages B, they do not allow deletion of referential 3rd person subjects, we should conclude that they, too, lack a D-feature in I. However, the prohibition against deleting referential 3rd person subjects seems to be but a special case of a more general condition prohibiting null subjects under almost any circumstances, in these languages. In (43a), for example, the conditions seem right for deletion of the embedded subject, as it has an antecedent in the next clause up, and furthermore the embedded I has the right features, yet a null subject here is sharply ungrammatical. In (43b) the subject is generic, yet cannot be null.

(43)  
a. Vous pensez [que *(vous) parlez bien Anglais]. (French)  
you.PL think-2PL that you.PL speak-2PL well English  
‘You think that you speak English well.’  
b. These days *(one) doesn’t ever need cash except on the bus.

Why are (43a,b) ungrammatical? A simple answer is that these languages have a stricter, ‘phonological’ EPP-condition which not only requires a filled specIP, but a pronounced specIP. Various objections can be raised against this hypothesis: For one thing, null subjects actually do occur, although under more restricted circumstances than in, for example, Finnish. A well-known case, found in many non-null-subject languages, is the 1SG null subject typical
of personal letters and diaries (see Haegeman 1990), but also not infrequently heard in spoken language.

(44) Can’t tell you how happy I am to see you.

Another case is ‘conjunction reduction’:

(45) John witnessed the accident, but (he) doesn’t want to talk about it.

That this may be correctly regarded as a case of a null subject (rather than, say, coordination of constituents smaller than IP) is suggested by the observation that there are non-null-subject languages which do not allow subject-drop even in this context; see Schmidt, Odden & Holmberg 2002: 13 for an example.

I will leave the proper account of non-null-subject languages in this unresolved state. See Holmberg 2003 for a discussion of Swedish.

12. Discourse pro-drop

A well-known division among null-subject languages is between those that rely on rich agreement, and those that have no agreement but rely exclusively on the wider discourse context for recovery of the null subject’s features. Examples of the latter type are Chinese, Japanese, Korean, Malayalam, and Thai. 25


      girl leave-ASP because tire-ASP

‘The girl(s) left, because she (they) were tired.’
I have argued that languages with subject agreement (Agr) cannot have a pro subject of the classical type, i.e. a nominal category which is inherently unspecified for number, person, and gender/class. Those languages need a specified subject to assign values to Agr. The agreementless languages do not have that problem, and may, on that account have a subject which is inherently unspecified. As long as the subject has features sufficient to support a theta-role and a case, and possibly check the EPP (insofar as the EPP is active in the language and some other category does not check it), it meets the needs of Narrow Syntax.

That is to say, pro exists, but (somewhat paradoxically, given the traditional view of pro) only in languages which do not have agreement. In a way this echoes Rizzi’s (1986) suggestion that discourse pro-drop languages like Chinese and Japanese are exempted from the licensing and identification conditions he proposed because their grammar does not employ $\phi$-features at all. I do not need to make a claim as strong as that; Instead, the crucial property of those languages would be that they have no unvalued $\phi$-features.\(^{26}\)

Another property which seems to be shared by the Agr-less discourse pro-drop languages is that they allow bare N arguments (see Tomioka 2003, Jayaseelan 1999). Tomioka (2003) argues that null arguments in Japanese are (or at least can be) bare Ns. Since pro is, by definition a bare N, a minimally specified nominal category, Tomioka’s arguments can be taken to support the claim that null arguments in Japanese are instances of pro.

13. Conclusions and some residual issues
The starting point was the incompatibility of the classical theory of *pro*, according to which *pro* is inherently unvalued and assigned feature values by Agr, with the feature theory of Chomsky 2000, 2001, in which Agr is inherently unvalued, and assigned feature values by the subject. Taking the latter to be right, two competing hypotheses were formulated:

*Hypothesis A:* There is no *pro* in null subject constructions. Instead Agr, the $\phi$-features of I, is itself an interpretable category, a referential, definite pronoun phonologically expressed as an affix.

*Hypothesis B:* There is a null subject in specIP in null subject constructions. The null subject is specified for interpretable $\phi$-features, and values the uninterpretable features of Agr just like any other subject does.

It was shown that Hypothesis B is right, at least in the case of definite and logophoric null subjects in Finnish: In definite or logophoric null subject constructions a pronominal subject checks the EPP and thereby excludes merge of an expletive, or movement of another category to specIP for EPP-reasons. The answer to the title question is thus that there is a proper null subject in specIP in finite null-subject sentences in Finnish. but it is not *pro* as defined in Chomsky 1982 or Rizzi 1986, but either a null pronoun which is specified for $\phi$-features but lacks D (the 3rd person null subject), or a fully specified pronoun with D, which is deleted in the Phonology (the 1st and 2nd person null subjects).

In Finnish the null $\phi$P pronoun must either be bound by a QP or a DP in the matrix clause or else be interpreted as generic. In the former case it must be in specIP, in the latter case in specvP. It cannot be interpreted as definite. This cluster of properties is found in a number of otherwise unrelated languages, called partial null-subject languages. These languages are distinguished from consistent null-subject languages (Spanish, Greek, Turkish, etc.) in that they lack a D-feature in I. Presence of D in I means that a D-less subject pronoun can be interpreted as definite. Absence of D in I means it must be bound or interpreted as generic.
Where does this leave the correlation between rich agreement and null subjects which undoubtedly exists, although there are counterexamples both ways (see Cole 2000, Huang 2000)? Clearly, in cases like (1) or (5) the subject’s feature values are identified by virtue of the agreement on the finite verb. However, this identification is not due to a rule or condition of (narrow) syntax, as in Rizzi 1986 and related work, but to sentence processing. Narrow syntax is oblivious to whether pronouns or inflectional affixes do or do not end up being pronounced. Sentence processing, by contrast, is obviously highly dependent on phonological features: If the subject is null, and the agreement on the finite verb is also null or not sufficiently distinct, then recovery of the features of the subject will fail, or will have to rely on information from the discourse.

How are the null subjects discussed above related to PRO, the null subject of non-finite clauses? As regards obligatorily controlled PRO there is an ongoing debate whether it is derived by movement; see Hornstein 1999, 2000, 2003, Landau 2003, Boeckx & Hornstein 2004. If it is, obligatorily controlled PRO is not a null pronoun but a copy. However, non-obligatorily controlled (NOC) PRO, as in (47), for example, is a null pronoun; see Hornstein 1999, 2000, 2003.

(47) John thinks that [PRO shaving himself] is important.

As non-finite clauses generally do not have Agr in I, PRO may, on that account, be a bare N, like pro in discourse pro-drop languages. Alternatively, if bare N arguments are not tolerated outside the discourse pro-drop languages even in the form of pro, then PRO may be a null \( \emptyset \)P. In that case, and since non-finite clauses do not have a D-feature in I, we expect PRO to have properties similar to the null 3\textsuperscript{rd} person subject in Finnish. This seems, indeed, to be the case. NOC PRO is interpreted as bound if there is a suitable DP binder. If not, it is interpreted as
generic. As noted by Landau (2003) NOC PRO is logophoric. The structural conditions on the binding relation are broadly similar to the ones that obtain for the null 3<sup>rd</sup> person subject in finite clauses in Finnish in that c-command is not a requirement, compare (48a) and (9c), repeated here as (49a), but the structural relation is not unrestricted, either, compare (48b) and (9f), repeated here as (49b). In (48a,b) the bound reading is forced by the reflexive, hence (48b) is ill-formed. Without the reflexive, the generic reading of PRO is available.

(48)  a. To John’s disappointment, [PRO shaving himself] was impossible.
    b.*John’s mother thinks that [PRO shaving himself] is important.

(49)  a. Se oli Tarjalle i pettymys [ettei hän /Ø /Ø j saanut lukea latinaa koulussa].
      it was Tarja-ALL disappointment that-not she/ Ø could study Latin school-INE
      ‘It was a disappointment for Tarja that she couldn’t study Latin at school.’

                      b. Se oli Tarjan äidille i pettymys [ettei hän /Ø /Ø j saanut lukea latinaa koulussa].
      it was Tarja’s mother-ALL disappointment that-not she/ Ø could study Latin school-INE
      ‘It was a disappointment to Tarja’s mother that she could not study Latin at school.’

A more detailed comparison of the conditions on the bound reading of the Finnish 3<sup>rd</sup> person null subject and that of PRO is left for future research.

References


Chomsky, Noam. 2001b. Beyond explanatory adequacy. Ms. Department of Linguistics and Philosophy, MIT.


Anders Holmberg

School of English Literature, Language, and Linguistics

University of Newcastle

Newcastle upon Tyne NE1 7RU

UK

anders.holmberg@newcastle.ac.uk
Research for this paper was supported by grant R280168 from the AHRB. Special thanks to the other members of the AHRB-funded project ‘Null subjects and the structure of parametric theory’: Ian Roberts, Dave Willis, Theresa Biberauer, Chris Johns and Michelle Sheehan. Numerous other people have contributed with data and/or comments on earlier versions of this paper, or on my talks and lectures on these issues. The following is a very incomplete list: Anna Cardinaletti, Melvyn Cole, Annabel Cormack, Mayumi Hosono, Christer Platzack, Hannu Reime, Peter Svenonius. Thanks also to two anonymous referees for LI, one of whom demanded a thorough revision of the paper, which helped to improve it considerably. All remaining errors are my own.


1 The focus in Rizzi 1986 is on a form of pro which is not licensed and identified by INFL, namely object pro in Italian, hence the very general formulation of the licensing condition; In Italian transitive V is an X°. Case-marking and licensing an object pro, which, as Italian V does not have φ-features, is interpreted with arbitrary reference.

2 The predicate in (7) can denote a plurality of distinct occurrences of being tired, but crucially it need not do so. Correspondingly, a verb with singular morphology can denote a plurality of events or states, for instance as in (i):

(i) La chica se cansaba todos los días
‘The girl became tired every day.’.

There are languages that have an interpretable plural form of the verb; see Newman 1990, Schmidt & al. 2002. This verb form is “used primarily to indicate plural action, either on the part of several agents /…/ or applied to several objects on one or several occasions, or an event that is taking place on several occasions. (Schmidt & al. 2002: 10). In these languages the plural (or ‘pluractional’) form of the predicate can only denote a plurality of events or states.

3 A pure expletive does not trigger agreement and appear not to be assigned Case. Their only function is to satisfy the EPP (see Chomsky 1995: 288).

4 It seems that the antecedent can have any syntactic function as long as it is the only possible antecedent in the next clause up, as in (9c,d). If there are several arguments in that clause, then a hierarchy of accessibility applies, as discussed in Gutman 2004, where the subject is the favoured antecedent. (i) shows that the antecedent has to be in the next clause up:

(i) Se oli Tarjalle pettymys [kun tuli selväksi [ettei hän/*Ø saanut lukea 

        it was Tarja-ALL disappointment when became clear that-not she could study 

Latin school-INE

‘It was a disappointment to Tarja when it became clear that she couldn’t study Latin at school.’

5 Here, as elsewhere in this paper, ‘EPP’ refers to the EPP-feature of I, not the generalized EPP-feature of Chomsky 2000. The discussion here presupposes that there is variation across languages regarding which categories can satisfy the EPP. See Holmberg 2000.

6 Vainikka & Levy (1999: 636) seem to say that a 1st or 2nd person subject will always raise to specIP (specAgrSP in their framework). This is not the case, however.
A referee for LI (henceforth R) questions whether constructions such as (19a,b) have any implications for the distribution of sitä in (21). R points out Icelandic has two homonymous expletives það, according to Thráinsson 1979. One, which Thráinsson called the ‘central’ expletive, is used in standard presentational/impersonal sentences. The other, called the ‘demonstrative’ expletive, is used in clefts, such as (i), and certain special expletive constructions.

(i) það er ég sem hef farið til Parisar.

EXP is I that has gone to Paris

R suggests that sitä in (19a,b) is a demonstrative expletive, like það in (i). In that case it has no implications for the distribution of purportedly central sitä in (21, 22), just like (i) has no implications for (ii).

(ii) *það tala ég ensku.

EXP speak I English

R then suggests that the expletives in (21, 22) and (ii) might be excluded because a central expletive can never scope over a referential subject, whether it is an XP in specIP, or Agr in I. Consequently, (21) provides no evidence that there is a null XP in specIP blocking out the expletive.

Finnish has two expletives, too, but they are morphologically distinct. The demonstrative expletive is se, the nominative form of ‘it’ (see (9c,f) and (11b)), the central one is sitä, the partitive form of ‘it’.

7
The argument in the text therefore stands: The central expletive sitä can scope over a referential subject, when that subject does not occupy specIP. Thus the best explanation of the ill-formedness of (21a,b) is that in those constructions the null subject checks the EPP.

8 See Holmberg 2001 for a detailed account of polarity focusing in Finnish. (24a) is a possible reply to the question in (i), optionally retaining just the finite verb:

(i) – Oletko sinäkin käynyt Pariisissa?

be-2SG-Q you-too visited Paris-INE

‘Have you been to Paris, too?’

– Olen (minäkin käynyt Pariisissa).

be-1SG I-too visited Paris-INE

‘Yes I have.’

As discussed in Holmberg (2001), the reply is derived by move of the finite verb/auxiliary to C followed by optional IP-deletion. A corresponding deletion is impossible in (ii), as expected if it is not derived by move of the auxiliary to C.

(ii) Sitä olen *(minäkin käynyt Pariisissa).

9 The correctness of the prediction has been questioned for Spanish (Suñer 2004) and for European Portuguese (Costa & Galves 200?). See also Cardinaletti 2004.
Matters are complicated by the fact that an object or adverbial may scramble to preverbal position when (a) the sentence is introduced by a focused category, and (b) the object or adverbial is not Main News. If the focused category is then not the subject, the result will be a sentence with three or more categories preceding finite I. Thus for instance (i) is well formed:

(i) Kukkia Mikko Annalle antoi.
   flowers Mikko Anna-ALL gave
   ‘It was flowers that Mikko gave to Anna.’

See Vilkuna 1995 and Holmberg 2001. In that paper I argue that the finite verb is, in fact, preceded by only two constituents even in this case: The category checking the EPP is a remnant VP containing the subject and (in this case) the indirect object.

The subject can be null in (29a); As predicted, the fronted locative is then necessarily contrastive, and cannot be embedded in a conditional sentence.

(i) Pariisissa olen känyt (mutten Roomassa).
   Paris-INE be-1SG visited but-not Rome-INE
   ‘I have been to PARIS, but not Rome.’

(ii) *Jos Pariisissa olisin känyt, ...
    if Paris-INE be-CON-1SG visited

As one referee notes, this looks like a special case of Burzio’s generalization. Its relation to Burzio’s generalization is, in fact, rather complicated, as discussed in particular by Nelson (1998). One complication is that the subject assigning values to Agr and the object assigned accusative may be in different clauses, as long as no finite clause-boundary separates them. Compare (i) and (ii). The object of the embedded clause is accusative in (i) because the main
clause contains Agr agreeing with a nominative subject, while it is nominative in (ii) because
the main clause does not contain an agreeing Agr.

(i) Minä menin Saksaan ostamaan auton/*auto.
    I went-1SG Germany-ILL buy-INF car-ACC/car-NOM
    ‘I went to Germany to buy a car.’

(ii) Minun täytyy mennä Saksaan ostamaan auto/*auton.
     I-GEN must go Germany-ILL buy-INF car-NOM/car-ACC

13 Crucial evidence that presence of a nominative subject alone is not a sufficient condition
for accusative case, in the absence of an agreeing Agr, is provided by (i):

(i) Me voidaan ostaa auto/*auton.
we-NOM can-PASS buy car-NOM/car-ACC
    ‘We can buy a car.’

In colloquial Finnish a 1PL active meaning is normally rendered by the passive/impersonal
form of the verb in construction with a 1PL nominative pronoun. The passive/impersonal verb
is inflected for tense and mood but not for agreement. As shown, the object has nominative,
not accusative, case (see Timberlake 1972, Reime 1993).

14 Compare, for example (i) and (ii), where a null subject is completely impossible in (ii).

(i) Jarille tuli selväksi ettei (hän) saisi palkintoa.
Jari-ALL became clear that-not he get-CON prize
   ‘It became clear to Jari that he wouldn’t get a prize.’

(ii) Kirjasta tuli selväksi ettei *(se) saisi palkintoa.
book-ELA became clear that-not it get-CON prize
   ‘It was clear from the book that it wouldn’t get a prize.’
They are structurally very similar, but in (ii) the embedded clause does not report the attitude or mental state of the argument in the matrix clause which binds the embedded subject.

The facts are the same if the argument in the higher clause is quantified: preposing an object or adverbial precludes binding of an embedded subject, leaving the generic reading as the only alternative. Thus (i) can only mean ‘Every student knows that the assignment can’t be solved’, not ‘Every student knows that he can’t solve the assignment’.

(i) Jokainen oppilas tietää ettei tehtävää pysty ratkaisemaan.
    every students knows that-not assignment can solve

Their typology is very similar to that of Cardinaletti & Starke (1999), discussed below.

The other criteria distinguishing a φP from as DP or NPin Dechaine and Wiltschko 2002 are (a) It does not have the internal structure of a NP nor that of a DP, and (b) it can be a predicate as well as an argument. Criterion (a) is inapplicable to a null pronoun. Criterion (b) is also inapplicable since in order to function as a predicate, a pronoun must be focused (as in The one who did it was HIM.), and a null pronoun cannot be focused.

The difference between our φP and D&W’s could be that the former is a NumberP while the latter is a PersonP. However, in the absence of a theory which would explain why Person rather than Number is crucial for a pronoun to support coreference, I prefer to leave it open which feature it is that distinguishes D&Ws φP from our φP.

I am indebted to Michelle Sheehan and Marc Modesto for the Brazilian Portuguese data, Aarti Nayudu for the Marathi data, and Ur Shlonsky for the Hebrew data.

They do have a 3PL null subject corresponding to arbitrary they in English, as in (i):

(i) In France they eat snails.

21 What about the putative null expletives or quasi-arguments in cases like (i) and (ii) (compare (11a,b))?

(i) Oli hauskaa että tulit käymään.
   was nice that came-2SG visiting
   ‘It was nice that you came by.’

(ii) Sataa vettä.
   rains water-PAR
   ‘It’s raining.’

Recall that the EPP in Finnish is activated only if there is a category in IP which can check it. In (i) and (ii) there is no such category. A clause cannot check the EPP in Finnish, nor can the partitive complement in (ii) do so.

(iii) *[Että tulit käymään] oli hauskaa.

(iv) *Vettä sataa.

As mentioned in Holmberg & Nikanne 2002, (iv) is considerably improved if another predicate is added:

(v) Vettä sataa kaatamalla.
   water rains pouring-ADE
   ‘It’s pouring down.’

Ignoring this complication, we may conclude that I in (i) and (ii) does not have an active EPP, and therefore the constructions have no null subject. As mentioned in section 6, an overt expletive is sometimes optional in such cases, which explains the optional pronoun in (11a).

22 The distinction between strong and deficient pronouns in C&S appears almost identical to that between DP and φP in D&W. But D&W’s NP pronoun type is not necessarily a clitic.
An interesting possibility is that Finnish and the other partial null-subject languages have a counterpart to D in I which functions as a generic operator binding the null \( \emptyset \)P. This could explain why the generic null pronoun must remain c-commanded by I.

Since K in the form of Agr in C&S licenses not just null pronouns but weak pronouns in general, their theory would seem to predicts that partial null-subject languages should lack not just null weak pronouns, but overt weak pronouns as well. More research is needed to establish whether the prediction is right. It could be noted that none of the partial null-subject languages discussed in Holmberg (to appear) have sentential clitic pronouns (the other type of deficient pronouns in Cardinaletti & Starke 1999).

Thanks to Nianling Yang for the Chinese examples.

There are also well known counterexamples to the generalization that Agr-less languages allow discourse pro-drop. The Mainland Scandinavian languages, for example, have no subject-verb agreement, yet do not allow discourse pro-drop (with exceptions noted in section 11). In present terms the question is: Why do Danish, Norwegian and Swedish not allow a pro subject, identified by a discourse antecedent, like Japanese, Korean, etc.? The simple answer (possibly simplistic, given the exceptions mentioned) is that Danish, Norwegian and Swedish have a phonological EPP-condition which rules out any kind of null subject.