

The functions of a high-frequency collocation in native and learner discourse: The case of French *c'est* and Swedish *det är*

INGE BARTNING AND BJÖRN HAMMARBERG

Abstract

*This cross-linguistic study investigates the functionality and use of one particular linguistic collocation in each of two languages, viz. the French *c'est* and the Swedish *det är*, both meaning 'it is'. The data are drawn from conversational speech production by adult native speakers and second language learners. The investigation shows that these collocations are highly polyfunctional as components in different grammatical constructions, that they play a significant role as a formula in the planning and execution of utterances and that they show a characteristic distribution of frequency of use and a characteristic profile of development in learner speech. A usage-based view of how language develops in individuals is adopted in the analysis of these findings, showing that these various properties are connected to each other and that they have an important bearing on the development of learners' interlanguages.*

1. Introduction

The role of specific high-frequency items in language use and acquisition has not yet been very much studied. However, it is becoming clear that the fact that an expression is frequently used may have important implications for the build-up of language competence. Recent research (e.g., Bybee and Hopper 2001a; Ellis 2002) has called attention to the effects that frequency of exposure and use has on the establishment of linguistic structure. According to a usage-based view of how language develops in individuals (see, e.g., Langacker 1999), the repeated use of specific linguistic items leads to their gradual entrenchment as cognitive routines in the speaker's mind.

On the other hand, frequency itself has a cause. There must be some motive or some factors causing a linguistic expression to be used with a certain degree of relative frequency. The need to communicate certain ideas, or the usefulness of putting messages in certain forms may constitute such factors. If the item in question is a grammatical construction, it may have functional

qualities which motivate its being used over and over again in formulating utterances. This suggests that a study of the connections between functionality and frequency may contribute to our understanding of the process of language acquisition. The linguistic functions of high-frequency items therefore call for investigation.

We will attempt this here in exploring the functionality and use of one particular linguistic collocation in each of two languages, viz. the French *c'est* and the Swedish *det är*, both meaning 'it is'. (We will use the term 'collocation' here in a general sense, referring to a regularly occurring sequence of words.) Our data will be drawn from conversational speech production by adult native speakers and second language learners of the two languages.

Predications of the form 'it is X' are very frequent in these two languages. The high frequency in spoken French of items such as *c'est* 'it is' and *il y a* 'there is/are' has been noted by several researchers (cf. Blanche-Benveniste 1990; Gadet 1989; Morel and Danon-Boileau 1998; for early French interlanguage, see Trévisse 1986, Véronique 1994). In Swedish, word counts based on spoken corpora show *det* 'it' to be by far the most frequent word in the language and *det är* 'it is' the most frequent two-word combination (Allwood 1999).

A striking fact is the polyfunctional nature of French *c'est* and Swedish *det är*. There is a high degree of parallelism between the two languages in the use of *c'est* and *det är*, as well as some interesting differences. This makes a crosslinguistic approach attractive in investigating the workings of *c'est/det är* in native and learner speech in the two languages. French and Swedish both belong to the West European typological area of *place-holder languages* (or languages which do not allow pro-drop, in Chomskyan terminology) (Beckman 1934; Hammarberg and Viberg 1977). This accounts for the abundant use of constructions with neutral and dummy pronouns. As we will discuss in greater detail below, French *c'est* and Swedish *det är* play a central role in structuring sentences and discourse in these languages and occur in a wide range of different contexts, in speech even more than in writing. A neutral pronoun such as French *ce* or Swedish *det* is often very useful in achieving simple reference to complex discourse entities, permitting vague or underspecified reference (cf., e.g., Fraurud 1992; Fraurud and Hellman 1999; Bartning 1997b), or serving as subject place-holders in different syntactic constructions.

The fact that *c'est* and *det är* can often be used in the initial part of sentences, as a component of a variety of constructions where they introduce sentence rhemes, makes them useful as a widely applicable formulation resource in the ongoing production of utterances. In particular, they are available to the speaker as a flexible means in the utterance-formulation process since there may be several alternative possibilities to complete the utterance after *c'est/det är*, and the syntactic choice of how to continue is often still open at that point. As we shall argue below, there is reason to assume that *c'est* and *det är* tend to

work as units, i.e., as formulaic sequences with a significant role in facilitating utterance processing (cf., e.g., Wray 1999; Wray and Perkins 2000; Forsberg 2006).

Previous studies of French *c'est* by Bartning (1997b, 2006) and the Swedish neutral pronoun *det* by Hammarberg (2000) indicate that constructions with *c'est/det är* are also readily adopted and used by second-language learners of French and Swedish. In the present contribution we will take these studies further and compare native usage with learner production at different stages of L2 competence. We will make use of two comprehensive corpora of recorded conversations, the French *InterFra Corpus* (Bartning 2002) and the Swedish *ASU Corpus* (Hammarberg 1999). We will attempt to explore the following main questions in some detail:

- (a) What is the grammatical distribution of *c'est* and *det är* in the speech of the native speakers and the learners of French and Swedish, respectively?
- (b) What is the quantitative distribution of the different uses of *c'est* and *det är* across native and non-native speakers and across different stages of interlanguage development? In what respects do native speakers and learners converge or diverge in their ways of making use of these items? Do longitudinal profiles of acquisition appear?
- (c) What role do the items *c'est* and *det är* play in the ongoing planning and execution of utterances in native speakers and learners?

2. Some crosslinguistic remarks on the uses of *c'est* and *det är*

Although the functions of French *c'est* and Swedish *det är* are similar to a great extent, there are also some important differences in distribution, mainly due to the different distribution of the pronominal elements, *ce* and *det*.

Both French *ce* and Swedish *det* are used as third person singular pronouns, either in a referring function or as dummy pronouns forming a part of various sentence constructions. What we focus on here is the role of these forms as 'neutral' pronouns, i.e., neutralising the number and gender variation that applies to third person pronouns in other contexts. We thus concentrate on cases where *ce* replaces subject pronouns *il/elle/ils/elles* and the invariable form *det* replaces *den/det/de*. A function of *ce* and *det* which we will leave aside here is that of a determiner, i.e., as an article or a demonstrative.

Compared with French *ce*, Swedish *det* has a wider and more varied distribution. While French pronominal *ce* occurs only in subject position with *être*, or qualified by a relative clause (*ce qui, ce que*), Swedish *det* occurs in different syntactic positions, as subject, predicative, object, or governed by prepositions, and it can combine with a variety of verbs. In the role of a neutral pronoun, *det* acts as a counterpart not only to *ce*, but also to *cela, ça, il* (in combinations such as *il est, il fait* and *il y a*) and the object pronoun form *le* in certain con-

texts (cf. for French, e.g., Olsson 1986; Riegel et al. 1994; Pedersen et al. 1982; for Swedish, e.g., Sundman 1980; Holmes and Hinchliffe 2003; Telemann et al. 1999; Hammarberg 2000). In the present study, we focus on a comparison of *c'est/det är* as the most frequent and functionally most varied collocation with the neutral pronoun *ce* and *det* in French and Swedish, respectively, which means that we concentrate on the combination with the verb *être/vara*, restrict ourselves to the present tense, and leave out possible occurrences of the plural *ce sont* as well as Swedish instances of verb-subject order (*är det*), or intervening elements (*det inte är*; *det kanske är* etc.).

In both languages, a prominent role for *c'est/det är* is to introduce a predication, since in most types of constructions, *c'est/det är* occurs in the opening part of sentences or clauses, i.e., either in absolute initial position, or preceded, for example, by conjunctions, subjunctions, discourse particles, interjections, or a predication of saying or thinking (such as French *et* 'and', *mais* 'but', *donc* 'so', *bon* 'well', *enfin* 'finally', *en fait* 'in fact', *alors* 'then', *ah oui* 'oh yes', *non* 'no', *je pense/trouve/crois que* 'I think that'; Swedish *och* 'and', *men* 'but', *så* 'so', *för* 'for', *fast* 'though', *att* 'that', *om* 'if', *ja* 'yes/well', *nej* 'no', *okej* 'okay', *jag tror/tycker* 'I think'; for the use of certain discourse particles in the InterFra corpus, see Hancock 2000). Especially in French, utterances very often contain a longer opening sequence, known as the *préambule* (cf. Morel and Danon-Boileau 1998, Conway 2005). *C'est* will then occur as the introducer *par excellence* of the rhematic part of the utterance.

3. Data for the present study

3.1. The French part

The French data used in this study originate from interviews with Swedish learners of French of the *InterFra Corpus*, which was compiled in the 1990s at the Department of French and Italian, Stockholm University (Bartning 1997a). This corpus contains speech productions (interviews, picture story narrations and video film retellings) and some written material collected from Swedish learners of French at five different levels and from native students.

The *InterFra Corpus* contains different subcorpora of 38 university students (hence called Uni, age 19–25) (6–8 students in a longitudinal group followed during four terms, 24 in three cross-sectional groups at three university levels, 6 teacher students in a longitudinal group before and after a term abroad), 20 students at secondary school with 3,5 years of French collected on two occasions with one year's interval (hence: Sec, age 16–17), 10 beginners followed during one to two terms (collected on 5 to 10 occasions) at the university (hence: Beg, age 19–26) and, as baseline data, 20 native speakers who were exchange students at Stockholm university, mostly from northern France with standard French as their dialect (NS, age 19–25). The collected material makes a total

of 500 000 word tokens (For a later description of the corpus, see Bartning and Schlyter 2004).

For this particular study we have chosen

- first term interviews with 8 of the university students with 4,5–6 years of French at secondary school. They have already passed one month at the university and have spent 3–12 months abroad in French-speaking countries.
- first interviews with 8 of the secondary school students after 3,5 years of French at secondary school, some of them with 1–2 weeks holidays in France;
- first interviews with 8 beginners who just started (three weeks ago) a beginners' course in French at the university, with no preceding experience and no formal training of French but with 1–2 weeks of holidays in France, mostly with Swedish friends; the developmental aspects of the *c'est*-structures are supported by recent longitudinal data from four of the beginners in 5 consecutive interviews during their first term
- interviews with 4 of the French native speakers

The interviews have been transcribed according to standard French orthography and a key of the transcription conventions is given in the Appendix.

3.2. *The Swedish part*

The Swedish data used here are part of the *ASU Corpus*, which was compiled in the 1990s at the Department of Linguistics, Stockholm University. This is a text corpus of Swedish conversational speech and written essays collected from foreign and native university students in Stockholm, totalling ca 500,000 words. A full description of the corpus is given in Hammarberg (1999). For the present purpose, the oral material from six learners and six native speakers has been used.

The learner part of the corpus is longitudinally designed, recording the developing learner language from the beginner stage when the informants had just arrived in Sweden and began the university language preparatory course, to a stage where they were participating in university studies in Swedish. The part used here comprises conversational speech from two speakers each of Chinese, Greek and Portuguese, one woman and five men, aged from 19 to 23. They all had prior L2 knowledge of English and one or two more languages. These learners can be broadly characterized as a '*semi-formal*', '*qualified*' and '*fast*' type of young adult L2 learners: '*semi-formal*' since they lived at the time in the Stockholm area and received their Swedish language input partly from the course and partly from the outside environment; '*qualified*' in the sense that they all had secondary education, previous experience with foreign languages, and strong instrumental motivation to learn the language of the country in order to proceed with the studies in their fields; and relatively '*fast*' since they advanced from the absolute beginner stage through or close to the proficiency

level required for university studies in about a year. The learners were recorded on ten occasions: five sessions at three week intervals between September and December 1990, four sessions at intervals of about a month from January to May 1991, and one session in March 1992. All corpus material was collected in separate sessions outside the language course and was not evaluated in the course. The oral corpus part consists of audio-taped conversations of about half an hour between one learner and one or two native speakers at a time, including different verbal activities such as interviews, discussions, picture story narrations, descriptions of objects, etc. The six learners' oral production totals ca 75,000 words, which yields a mean text length per learner of about 12,500 words.

The native part of the corpus was compiled after the completion of the learner part, to serve as target language control data. It was constructed along the same lines, except that it does not reflect a development over time. Efforts were made to make the native part as comparable as possible to the learner part, in terms of type of informants, contents and methods of data collection. Thus, a series of five half-hour sessions with similar contents was recorded with each informant. The native Swedish informants were four women and two men aged from 20 to 29, studying various non-linguistic subjects at the undergraduate level. They spoke Central Standard Swedish as heard in the Stockholm region (although one informant, the one called Z5, had a trace of a northern accent). The six native informants spoke ca 77,000 words in total, or about 12,800 words each on average.

The learners are referred to in the corpus as C1, C2, G2, G3, Q1 and Q2 (C for L1 Chinese, G for L1 Greek, and Q for L1 Portuguese speakers). The native Swedish informants are referred to as Z1 to Z6.

The recorded dialogues have been transcribed using a modified form of standard Swedish orthography in which the common 'spoken forms' are rendered. Thus, e.g., the form *de e* occurs for standard written *det är*. A key to the transcription conventions is given in the Appendix.

4. *C'est* and *det är* in different contexts in native and learner speech

This section will give a comparative overview of the use of French *c'est* and Swedish *det är* by native speakers and L2 learners. We will first (in Section 4.1) establish a taxonomy of contexts for *c'est* and *det är*, which will then (in 4.2) form the basis for a quantitative account of the various uses of *c'est/det är* by native speakers and learners in the French and Swedish corpora.

4.1. *A taxonomy of contexts*

The various structures which involve the use of *c'est/det är* differ in various ways with respect to discourse function and syntactic form. For the present

purpose, we will sort them into three groups, viz.,

- (A) cases in which the pronoun (*ce/det*) has some kind of anaphoric or deictic reference (structures 1–11 below);
- (B) cases in which the pronoun is either non-referring or refers ahead in the utterance (structures 12–16 below); and
- (C) cases in which *c'est/det är* is followed by some form of syntactic break (structures 17–19 below).

Before commenting on the characteristics of these main groups, we will present an inventory of structure types with selected examples from the French and Swedish corpus.

The examples are given here in the transcription in which they occur in the two corpora. Source references for the French examples are given as follows: speaker identification (group and pseudonym: Beg: Carin), task and chronological number of the task: Int(erview) 1.

Source references for the Swedish examples are given in the form of a speaker and session identification, e.g., Z4-1 for speaker Z4, session 1, C2-6 for speaker C2, session 6. As mentioned above, Z refers to native Swedish informants, and C, G and Q to learners with Chinese, Greek and Portuguese, respectively, as L1. The session number for the learners (ranging from 1 to 10) gives a hint about the current stage of proficiency.

A. Anaphoric or deictic *ce/det*

Structure 1. *C'est/det är* + Noun Phrase

- (1) a. *mais sinon c'est un pays très mélangé donc enfin.*
'but otherwise **it's** a very varied country after all' (NS: Françoise)
- b. *c'est un peuple qui est tellement chaleureux (I:mm) vraiment.*
'**these are** people who really are warm (generous)' (Uni: Laura Int1)
- c. *de e saker som ja funderar ganska mycke över.*
'**those are** things that I think quite a lot about' (Z4-1)
- d. *de e en nyhet ti mej (E: JA) tror ja.*
'**that's** news to me (E: YES) I think.' (C2-6)

C'est/det är is followed by an NP, simple or complex, definite or indefinite, with a noun or pronominalised. French *ce* functions as subject here. This is mostly also the case with Swedish *det*, although it may also occur as a preposed predicative, as in *nä de e han inte* 'no, (that) he is not' (Z4-3), *de e de* '(that/so) it is' (passim). In the quantitative tables below (Sections 4.2.1–2) we count the instances where a relative clause is embedded in the NP, as in examples (1b and 1c), as a separate subtype.

Structure 2. *C'est/det är* + Adjective Phrase

- (2) a. (Topic: How to get along with Swedes compared to Brazilians)
avec les Suédois c'est un peu difficile le contact [...] parfois [...] c'est plus difficile qu'avec les # c'est pas latin [=les Suédois ne sont pas latins](RIRE) mais euh il y en a beaucoup très très gentils. (I:mm)
 'it's more difficult with the # **it's** (= the Swedes are) not Latin (LAUGHTER) but there are many who are very very nice' (NS: Françoise)
- b. (Topic: Le Musée d'Orsay)
eh c'est c'est bon mais c'est st c'est c'est grande ma je ⟨ja⟩.
 'euh it's **it's** good but it's it's **it's** big ma I ⟨ja⟩' (Beg: Emelie Int1)
- c. (Topic: Anthrax bacteria can remain active for years)
å de e ju lite otrevligt kanske.
 'and **that's** a bit unpleasant perhaps.' (Z1-4)
- d. (Topic: How to shoot a film)
% de e inte så lätt.
 '% **it's** not so easy.' (G2-4)

C'est/det är occurs in a clause with a predicative adjective.

Structure 3. *C'est/det är* + Adverbial Phrase/Prepositional Phrase

- (3) a. *c'est euh en France justement.*
 '**it's** euh in France yes' (NS: Anne)
- b. *non pas le nom mais c'est c'est dans le même genre là.*
 'no not the name but it's **it's** of the same sort' (Uni: Marie Int1)
- c. *de e så himla längesen.*
 '**it's** so terribly long ago' (Z6-5)
- d. *de e på kvällen*
 '**it's** in the evening' (C1-3)

C'est/det är introduces an adverbial of place, time, manner etc., which can also have the form of a prepositional phrase.

Structure 4. *C'est/det är* + Infinitive Phrase

- (4) a. I: *quel est ton loisir préféré ?*
 E: *je crois que c'est lire.*
 'what is your preferred leisure time occupation? I think that it's to read' (Uni: Lena Int1)
- b. *non c'est eh li eh lire eh c'est ⟨lättare⟩.*
 'no it's eh li eh read eh it's easier' (Beg: Jan Int2)

- c. *nu finns de ju nästan bara en väg ut. å de e ju å anfalla.*
‘now there’s almost only one way out, y’know. and **that’s** to attack.’ (Z2-3)
- d. (Topic: Whether the informant experiences difficulties in writing Swedish)
... för ja läser teknik. (B: JA) så de e bara lösa problem. (B: JAHA MH) s- man behöver inte skriva uppsats.
‘... because I study technology. (B: YES) so **it’s** only [to] solve problems. (B: YEAH WELL) s- you needn’t write essays’ (C2-10)

C’est/det är is followed by an infinitive phrase. No instances occur with the native French speakers; it seems that French speakers would tend to choose other formulations here in oral language (cf. however some literary examples in Olsson 1986: 137: *C’est insulter la Fortune que de ne pas y croire* (Sagan)). Native Swedish examples (like (4c)) are few but normal for standard spoken Swedish. Examples like (4a), (4b) and (4d) suggest that this construction may be used by learners as a simplified structure, avoiding a more cumbersome, careful formulation.

Structure 5. *C’est/det är* + Subclause

- (5) a. *c’est parce que j’ai trouvé qu’il y avait un côté comment dire.*
‘**it’s** because I’ve found that there was something how to say’ (NS: Anne)
- b. *oui mais c’est parce que je ne peux pas rouler les+ les+ r.*
‘yes but **it’s** because I can’t pronounce the the -r’ (Uni: Laura Int1)
- c. *de e ju bara va ja tror att de heter.*
‘**it’s** only what I think it’s called’ (Z2-5)
- d. *de har ett hål. - - - de e för att man kan hänga de på en krok*
‘it has a hole. - - - **it’s** so that you can hang it on a hook’ (Q1-7)

The subclause may be introduced by all sorts of subjunctions. In the French corpus a very frequent one is *parce que*, which has been found to be an overused connector in learner discourse (statistically significant, Hancock 2000).

Structure 6. *Det är* + Stranded Preposition (only in Swedish)

- (6) *de e ja helt övertygad om.*
‘**that I’m** quite convinced about’ (Z1-3)

An anaphoric pronoun governed by a preposition is fronted, leaving the preposition stranded. The Swedish Verb-second constraint accounts for the word order *det är* in such cases. The initial placement (topicalisation) of a non-subject

pronoun is considerably more frequent in Swedish than, e.g., in English with its stricter adherence to SVO order.

Stranded preposition constructions do not occur in French.

Structure 7. ‘Tough’ construction (*c’est/det är* + Adjective + Infinitive Complement)

- (7) a. (Topic: Racism in France)
peut-être que je ressens pas [...] c’est difficile à percevoir.
 ‘maybe that I don’t feel [...] **it’s** difficult to perceive’ (NS: Française)
- b. (Topic: Discussion of musical genres)
c’est difficile à expliquer SIM (SIGH).
 ‘**it’s** difficult to explain’ (Sec: Tomas Int1)
- c. (Topic: What is going on in a picture)
de e hemskt svårt å säga.
 ‘**it’s** terribly difficult to tell’ (Z5-5)
- d. (Topic: Describing a handle)
de e lätt att = att hålla.
 ‘**it’s** easy to = to hold.’ (Q1-6)

This is the use of *ce/det* in the so-called ‘tough’ construction, which has been analysed as the raising to subject of an object of an infinitive phrase (see for French Pedersen et al. 1982: 325; for Swedish Sundman 1987: 114 ff.); for example in (7a): (i) *percevoir cela est difficile* → (ii) *c’/cela/ceci est difficile à percevoir*; (7c): (i) *å säga de e hemskt svårt* → (ii) *de e hemskt svårt å säga*. In the resulting construction the infinitive functions as a complement to the adjective: *difficile à percevoir*; *svårt å säga*. The infinitive then specifies the respect in which the adjective is relevant. This construction is possible with a certain set of adjectives. The majority of instances in our corpora, both from native speakers and learners in both languages, contain the adjective ‘difficult’: *c’est difficile à ... / de e svårt å ...*

We have chosen to categorise these cases separately from the ‘*c’est/det är* + AdjP’ cases in 2 above. In the French corpus there are very few examples of this type both in NS and in NNS production.

Structure 8. Object *det* raised from Subclause position (only in Swedish)

- (8) *de e ja inte säker på att afrikanerna tycker*
 ‘**that I’m** not sure that the Africans think’
 (= ‘I’m not sure that the Africans think so’) (Z4-1)

A fronted object pronoun in combination with the Verb-second constraint causes the word order *det är*. Cf. the comment on Structure type 6 above. An analogous construction with *c’est* does not occur in French.

Structure 9. *C'est/det är* with Left Dislocation

- (9) a. (Topic: Comparison between Swedish and French salaries)
la moyenne euh des salaires c'est plus élevé que # les gens vivent mieux en général.
 'the mean salary euh **it's** higher than # people have a better living in general' (NS: Françoise)
- b. *et euh mon âge c'est dix-neuf NON c'est vingt maintenant (RIRE).*
 'my age it's nineteen no **it's** twenty now' (Uni: Eva Int1)
- c. *astrologi de e ju en ny ny sorts religion eller va man ska säga.*
 'astrology **that's** a new new sort of religion or what you may call it' (Z2-5)
- d. *femti år de e inte så kul att fylla.*
 'fifty years **that's** not so fun to become' (Q1-7)

In both languages, left-dislocated structures with *c'est/det är* are quite frequent in colloquial speech. A thematic element of the sentence has been extracted to the left and is resumed by the pronoun *ce/det*. This dislocated element may be an NP, an infinitive, or a PP which occurs in direct connection with the following rhematic part introduced by *c'est/det är*. Especially in French, but sometimes also in Swedish, the left-dislocated part may contain more than one constituent; this is known in French grammar as a *préambule* (Morel and Danon-Boileau 1998; Conway 2005).

The part of the sentence which is introduced by *c'est/det är* coincides in form with a corresponding structure type under 1 to 8 above. However, we count the sentences with left dislocation here as a separate structure type.

Structure 10. *C'est/det är* in Pseudo-Cleft construction

- (10) a. (Topic: Sweden)
ce que j'aime bien ici c'est que les gens sont tous en haut.
 'what I like here **that's** that people are all at the top.' (NS: Françoise)
- b. *ce ce qui me manque c'est que je je n'ai pas été en France pendant deux ans.*
 'what I miss is **that's** that I haven't been in France for two years' (Uni:Lena Int3)
- c. *de som e svårast å måla de e typ ansikten å händer å sånt.*
 'what's most difficult to paint **that's** like faces and hands and such.' (Z4-4)
- d. (Topic: The meaning of the notion "idiot" in Greek)
de som inte e sta- statligt eller offentligt de e idiotiskt.
 'what isn't gov- governmental or public **that's** idiotic.' (G3-10)

We gather here pseudo-cleft sentences with a resumptive *ce/det*, which is a characteristic feature that makes this variant of pseudo-cleft a distinctive structure-type in French and Swedish. Actually, this is a special case of left dislocation, akin to our Category 9 above, in which the dislocated part consists of a pronoun modified by a relative clause (French *ce que X*; *ce qui X*; Swedish *det som X*; *vad som X*). The first element of the sentence is thus a periphrastic relative and the second, introduced by *c'est/det är*, is a sequence (NP, infinitive, *that*-clause) which maintains a complement relation with the verb of the relative (see Riegel et al. 1994: 432–433).

Pseudo-cleft sentences with a resumptive *ce/det* are characteristic of colloquial spoken French and Swedish. Sentences like (10a–d) would also be possible (although probably less frequent) without a resumptive pronoun: *ce qui me manque est X*; *de som e svårast å måla e X*. That variant, however, falls outside the scope of the present study.

Structure 11. *Det är* in terminal position in Subclause

- (11) a. *nu kan man till å me se hur långt de e / tjugi centimeter långt.*
 ‘now you can even see how long **it is** / twenty centimetres long.’
 (Z3-4)
- b. *ja vet inte riktigt hur lång tid de e.*
 ‘I don’t quite know how long time **it is**.’ (Q2-10)
- c. *bon c’est des bien grands mots pour ce que c’est mais*
 ‘well that’s pretty big words for what **it is** but’ (NS: Mélanie)

Here *c'est/det är* occurs in a subordinate clause. Unlike the other structure types in this taxonomy, *c'est/det är* does not function here as an introducer of a predication, but occurs toward the end of the sentence, to complete an ongoing utterance which is usually already planned in final shape by the speaker at this stage.

In our French corpus we have only occasional examples of Structure 11, such as (11c).

B. Non-referring or cataphoric *ce/det*

Structure 12. Impersonal

- (12) a. *hon pekar på klockan å visar att de e sent.*
 ‘she points to the watch and shows that **it’s** late.’ (Z3-3)
- b. *de e för varmt där.*
 ‘**it’s** too hot there’ (Q1-10)
- c. *de e dags att kliva upp.*
 ‘**it’s** time to get up.’ (Z5-5)

- d. *mm oui c'est froid à l'hiver.*
 'mm yes **it's** cold in winter' (Beg: Jan Int3)

In Swedish, a non-referring neutral *det* is required as a dummy subject pronoun with impersonal predicates of various types (cf. Teleman et al. 1999: Vol. 4, 58 ff.). Many different verbs are used with impersonal *det*. *Det är* occurs with adjectives (12a-b), participles or adverbials (e.g., *dags* in (12c)).

In standard French, impersonal *il est*, *il fait* are used in corresponding constructions, e.g., *il est tard* 'it is late', *il fait froid* 'it is cold'. We have, however, also found *c'est* used in this function by early learners in our corpus, as in (12d).

Structure 13. Existential

- (13) a. *de e ju fortfarande barn som blir misshandlade*
 'there are still children who get maltreated' (Z2-4)
- b. *de e skillnader mellan generation.*
 'there are differences between generations.' (C1-10)
- c. I: *il y a combien de pièces ? une pièce ? deux pièces ?*
 E: *oui. c'est // eh quatre pièces.*
 'there are // euh four rooms' (Beg: Jan Int4)
- d. *mais ça dépend si c'est quelque chose qui m'intéresse.*
 'but that depends if **there's** something that interests me' (Uni: Yvo Int1)

Swedish *det* is used as a dummy subject in existential sentences. These are the cases where English uses *there*. In these cases, standard French uses constructions with *il*: *il y a* 'there is/are'. However, the InterFra corpus evidences a typical learner use of *c'est* for *il y a* from the early stages to more advanced, as in (13c–d). It is interesting to note that *il y a* emerges later than *c'est* in the early stages of the InterFra corpus.

Sundman (1980), in her investigation of Swedish existential sentences and related constructions, lists the following formal criteria for identifying the existential construction: (i) A dummy *det* in subject position, (ii) an NP as logical subject, (iii) no predicative. She also points out (1980: 19 f.) that the logical subject occurs in object position, that the logical subject is indefinite, that the logical subject can in most cases replace *det*, and that a place adverb *här*, *där* 'here', 'there' can sometimes serve instead of a dummy subject as placeholder.

The function of this construction type is to express the logical subject as rhematic. It is typically used to introduce new referents in the discourse (Ekerot 1979: 96 f.). If the logical subject includes a relative clause, as in (13a), this contains new information. This distinguishes this type from the cleft construction which we categorise as Type 16 below. Instances where the logical subject

14 Inge Bartning and Björn Hammarberg

includes a relative clause, such as (13a and 13d), are counted as a separate subtype 13:2 in the quantitative tables below.

Structure 14. Anticipating

- (14) a. *pour un noir ou un magrhebin c'est beaucoup plus difficile de s'intégrer.*
'for someone from the Maghreb **it's** more difficult to get integrated' (NS: Anne)
- b. *parce que c'est pas naturel de parler en classe je pense.*
'cause **it's** not natural to speak in class I think.' (Uni: Yvonne Int1)
- c. *de e klart att de e bra å kunna andra språk också.*
'of course **it's** good to know other languages as well.' (Z3-5)
- d. *för de de e viktigt hur man presentera ett projekt*
'cause **it's** important how you present a project' (Q1-10)

In this type of construction *c'est/det är* is followed by a predicative adjective, and *ce/det* anticipates an extraposed constituent which constitutes the logical subject of the sentence. The extraposed constituent may have the form of an infinitive or a clause.

The extraposed constituent can replace *ce/det*; cf. (14b) *Parler en classe n'est pas naturel*. In written French the use of *ce* is often replaced by *il*: *C'il est permis de fumer* 'It is allowed to smoke'.

Ce/det functions like a cataphoric pronoun in these structures (in contrast to the *ce/det* in structure Types 4 and 5 above) in that it anticipates the extraposed logical subject. However, in this role it acts as a dummy, or expletive subject and is mostly discussed in these terms in the grammatical literature.

Structure 15. Right dislocation

- (15) a. *c'est trois ans la licence.*
'**it's** (= it takes) three years the bachelor's degree' (NS: Mélanie)
- b. *je trouve que c'est beaucoup plus intéressant la le province.*
'I think that **it's** much more interesting the countryside' (Uni: Yvonne Int1)
- c. *de e ju så himla omfattande de där.*
'**it's** so awfully big that' (= it's such an awfully big question, that') (Z6-5)
- d. *ja tycker de e så sjukt just de här att dom tillverkar ruskia vapen*
'I think **it's** so sick just the fact that they produce horrible weapons' (Z3-3)

The subject of a predicative sentence has been extracted to the right and is replaced in its canonical position by *ce/det*. Whereas French uses the neutral *ce* invariably in this case, the Swedish pronoun has to agree in number and gender with a dislocated noun, which causes *det* to alternate with other forms, e.g., *den är fin, den här tavlan* ‘it’s fine, this picture’ (Singular Uter).

Structure 16. Cleft

- (16) a. (Topic: Socially obedient)
je trouve que les Suédois sont beaucoup plus ou c’est peut-être les Français qui le sont beaucoup moins (RIRE).
 ‘I think that Swedes are much more or **it’s** maybe the French who are so to a lesser extent’ (NS: Anne)
- b. *c’est pas moi qui va investir.*
 ‘**It’s** not me who will make an investment’ (Uni: Sabina Int1)
- c. *de e i köket man använder den.*
 ‘**it’s** in the kitchen you use it’ (Z1-2)
- d. *ja ja kan tänka mej att de e därför som han fortfarande lever.*
 ‘well I can imagine that **it’s** therefore that (= that’s why) he is still alive’ (Z1-3)

French and Swedish use *ce* and *det*, respectively, as dummy pronouns in the cleft construction. The usual type of cleft sentences in French and Swedish has the form *c’est/det är* + NP/Adv/PP/V + Relative clause. French and Swedish clefts are formed (like clefts in English) by extracting a constituent from the uncleft sentence, introducing it by *c’est/det är*, and turning the remaining part of the sentence into a relative clause, e.g.,

j’aime Nicolas → c’est Nicolas que j’aime
 ‘I love Nicolas → **it’s** Nicolas that I like’

Cleft sentences usually function as emphatic structures, emphasising the element introduced by *c’est/det är*. In Swedish, this construction may in some cases coincide syntactically with existential sentences containing a relative clause (cf. Sundman 1980: 54 ff.). Thus, e.g., a sentence like (13a) above, if read in isolation, has two readings:

- (α) *de e ju fortfarande barn som blir misshandlade*
 ‘**there are** (= exist) still children who get maltreated’ (Type 13: existential)
- (β) *de e ju fortfarande barn som blir misshandlade*
 ‘**it’s** still children (= children are still the ones) who get maltreated’ (Type 16: emphatic cleft)

In (β), unlike (α), the relative clause is thematic and contains known information, and the sentence prosody differs, placing emphatic stress on *barn*. While the rhematic constituent is always indefinite in Type (α), it may be definite or indefinite in the emphatic clefts of Type (β).

In standard French, *c'est* is used in Type (β) and *il y a* in Type (α); however, as we showed in (13d) above, we have found occasional instances where the Swedish learners of French have extended the use of *c'est* to the existential function.

C. Formulation breaks

Breaks in the fluent formulation of the utterance are a natural part of the production of utterances. This is typical of impromptu speech, which is by definition spontaneous and not always fully planned before execution in a typical conversational situation. Such breaks may affect the syntactic structure of the sentence in various ways.

Our concern here are breaks which occur after *c'est/det är*, which is a point in the ongoing utterance where breaks typically tend to occur. In Swedish, *det är* is often followed by an unstressed modal adverb, such as *ju* 'you know', *väl* 'isn't it', or the negation *inte* 'not' etc. in the same prosodic unit before a break. In French, too, *c'est* is occasionally followed by the negation *pas* before a break. These cases are also included here. Taking into consideration what follows after the break, we distinguish cases of *repetition*, *reformulation* and *abandonment*, types which range from slight to more radical alteration of the current utterance. These phenomena offer some insight into the utterance-production process in that they bear witness of points of hesitation or replanning. The mere occurrence of a pause, filled or unfilled, at the location after *c'est/det är* (+ModAdv) has not been counted, however, although this is also clearly a hesitation phenomenon. It seems that the places in the utterance where pauses occur tend to vary much more and are not quite so typically connected with the item *c'est/det är*.

C. Structure 17. *C'est/det är* repeated

- (17) a. *mais bon c'est c'est c'est intéressant parce que c'est assez assez complet en marketing.*
 'but well **it's it's** it's interesting because it's rather rather complete in marketing' (NS: Françoise)
- b. *eh c'est c'est bon mais c'est c'est c'est grande ma je ⟨ja⟩.*
 'euh **it's** it's good but **it's it's** it's big ma I ⟨ja⟩' (Beg: Emelie Int1)

- c. *ja de e ju = de e ju ganska självklart i såna situationer på nå sätt.*
 ‘well **it’s** = it’s rather self-evident in such situations in a way isn’t it’ (Z1-4)
- d. *nej de e de e bra att åka bort tycker ja.*
 ‘no **it’s** it’s good to go away I think.’ (Q1-10)

C’est/det är (+ModAdv) is repeated (with or without an intervening pause) before the utterance is continued. The break can be interpreted as a moment of continued planning of a not yet fully planned utterance. Although no reformulation is observable in the uttered sentence in this case, some latent change of utterance plan may of course have taken place. Note that it is the fragment *before* the break that is counted here; the post-break part starting with *c’est/det är* is categorised separately in the appropriate category under A or B above.

Structure 18. *C’est/det är* + reformulation

- (18) a. *je veux dire c’est / # j’ai vu en en séminaire.*
 ‘let’s say **it’s** / # I saw at the seminar’ (NS: Mélanie)
- b. *eh c’est pas # eh mon eh st (instrument) est [...]*
 ‘it’s not # my instrument is [...].’ (Beg: Jan Int4)
- c. *ja tycker att de e / på nå sätt / ur moralisk synvinkel så e de väl ganska så förkastligt*
 ‘I think that **it’s** / in a way / from a moral point of view it’s rather reprehensible isn’t it’ (Z1-3)
- d. *så / = ja de e / ibland = e de = orättvist tror ja.*
 ‘so / = well **it’s** / sometimes = it’s = unfair I think.’ (C2-9)

In this case a restructuring of the ongoing utterance is obvious. (The restructuring breaks are rendered by “#” in the French and “/” in the Swedish transcripts.)

Structure 19. *C’est/det är* + abandonment

- (19) a. *et le club à cinquante SIM cinquante couronnes là c’est*
 ‘and the club (which costs) fifty SIM fifty crowns you know **it’s** (NS: PN)
- b. *mais en classe quand on va demander quelque chose tout le monde demande en suédois + et pour moi c’est*
 ‘but in class when you ask something everybody asks in Swedish and for me **it’s**’ (Uni: Yvonne Int1)
- c. *ja. de beror ju på / man kan ju argumentera för å emot i dom flesta frågor så att de e /*
 ‘well, it depends / of course you can argue for and against in most issues so that **it’s** /’ (Z1-4)

- d. *ja dom bara gå hemifrån å låter alla ljus på. å (B: JA) de e /*
 ‘yes they only go from home and leave all lights on. and (B:
 YES) it’s /’ (Q1-9)

Here *c’est/det är* (+ ModAdv) occurs in a position to introduce a predication, but the continuation is abandoned. The speaker starts an utterance with the common introducer *c’est/det är*, but fails to produce either a relevant message or its appropriate formulation. Nor is the utterance recast, as in Type 18 above; the fragment *c’est/det är* is left hanging. In many instances, this *c’est/det är* will serve as a turn-ending signal.

In summary, this taxonomy of contexts shows that French *c’est* and Swedish *det är* can participate in a wide range of different sentence structures in the respective language, especially in the function of introducing a predication. As we have seen, most of these various structures are produced by both native speakers and learners.

In the structures under A above (Structures 1–11), the pronoun refers to something given in the preceding discourse or in the situation. Especially the fact that the neutral pronoun can make simple reference to complex contents without unnecessary explicitness makes it widely useful in communication. One may expect that second language learners, once they discover this convenient mechanism in the language, are ready to use it freely and perhaps overuse it in a simplifying way in their attempts to express themselves in conversation.

The establishing of anaphoric reference in discourse has been discussed by several authors from a point of view which may help elucidate this function of the neutral pronoun. In general, we adopt the view of Brown and Yule (1983: 206 ff.) that the role of reference is to establish an entity in the speaker’s and hearer’s mental ‘discourse representation’ (or ‘discourse model’). The entity, or referent, may be some object which is represented by an antecedent noun, or it may be an event, a fact or a proposition, or some more complex information derivable from the foregoing discourse (Fraurud 1992). Following Fraurud and Hellman (1999), anaphora can be seen as “the retrieval of a *discourse referent* established into a (mental or computational) *discourse model* by means of a preceding expression” (Fraurud and Hellman 1999, italics in original). In some cases the referent is explicitly identified by, e.g., an antecedent noun, an infinitive phrase or a sentence. In other cases, the information referred to is not signalled explicitly in the preceding discourse, but has to be inferred when it is brought to attention by the anaphoric pronoun. That is, something in the discourse is established as an entity only when referred to by an anaphor. According to Fraurud and Hellman, this “involves hypostatization, the process by which something which is not an entity comes to be treated like an entity, resulting in what we here call *hypostatized discourse referents*” (Fraurud and

Hellman 1999, italics in original). A characteristic use of the neutral pronoun *ce/det* is to function as a *hypostatization anaphor* (term after Fraurud 2000) to establish mental representations of complex hypostatized entities from the preceding discourse. Consider Example (20) from the native Swedish informant Z2:

(20) (Topic: Watching a picture of an African storyteller in a group of listening people)

E: *va får du för associationer när du ser den där bilden? du kanske tänker på paralleller.*

‘what associations do you get when you see that picture? You may perhaps think of parallels.’

I: *en dagisgrupp tänker ja på.*

‘I think of a kindergarten group.’

E: *jaha.*

‘oh.’

I: *de e dom enda tillfällena tycker ja som som man samlar en grupp människor å berättar en historia.*

‘those are the only occasions I think when you gather a group of people and tell a story.’ (Z2-5)

Here, *de* relates to the foregoing *dagisgrupp* ‘kindergarten group’, but does not refer to this type of group as such. Rather, the event or occasion when such a group is assembled is hypostatized as a referent. By connecting *de* to *tillfällena* ‘occasions’ in the sentence that follows, the speaker makes this intended reference more explicit and accessible to the listener.

An important feature with anaphoric reference is that it may be more or less heavily *underspecified*, i.e., the entity referred to may be only vaguely identified by the anaphor. Underspecified reference is successful if the listener obtains a discourse representation which matches that of the speaker sufficiently well. This possibility of varying and reducing the precision of reference is clearly a useful mechanism in the language. On the one hand, the context may help indicate a referent sufficiently; on the other hand, great precision is often not necessary or even desirable. (21) is an example:

(21) *han % hyra % en del av v- villa me dom andra. och % vi laga mat tillsammans. % de e rolit tror ja.*

‘he rent[ed] a part of [a] house with the others. and we cook[ed] food together. **that is** [/was] fun I think.’ (C1-6)

Here the thing that is ‘fun’ may be interpreted variably as ‘to cook food’, ‘to cook food together’, ‘that we cooked food together’, or even ‘that we cooked food together in his house’. But precise reference is not important here, where

the main point with the utterance *de e rolit tror ja* is to express an attitude to the event in question.

Using *ce/det* to substitute for more careful reference has the advantage of simplicity and convenience. Another illustrative example of this vague, simple reference even in native speaker speech is (2a) above, which we repeat here for convenience:

- (2) a. (Topic: How to get along with Swedes compared to Brazilians)
avec les Suédois c'est un peu difficile le contact [...] parfois [...] c'est plus difficile qu'avec les # c'est pas latin [= les Suédois ne sont pas latins] (RIRE) mais euh il y en a beaucoup très très gentils. (I:mm)
 'it's more difficult with the # **it's** (= the Swedes are) not Latin (LAUGHTER) but there are many who are very very nice' (NS: Françoise)

For learners, the *ce/det* substitute may also be a means to avoid difficult morphology or uncomfortable sentence formulation as illustrated in (22)–(24) below:

- (22) *la femme c'est content.* (= la femme (elle) est contente)
 'the woman **it's** satisfied (= the woman (she) is satisfied)' (Beg: Carin, BD)
- (23) *les mots c'est eh très très difficile.*
 (= les mots (ils) sont très très difficiles)
 'the words **it's** euh very very difficult (=the words they are . . .)' (Beg: Carin Int1)
- (24) I: *qu'est-ce que tu prends comme moyen de transport pour aller à Stockholm ?*
 'what do you take as transport means to go to Stockholm?'
 E: *ah c'est Roslagsbanan* [= je prends Roslagsbanan or c'est R. que je prends]. *c'est un vieux train / un vieil train (I:mm).*
 'ah **it's** Roslagsbanan [=I take Roslagsbanan / It's R. that I take] it's an old train / an old train' (Uni: Eva Int 1)

In the structures under B (Structures 12–16) where *ce/det* is either non-referential or cataphoric, a different mechanism does the job. The entity spoken about is made clear in the coming utterance, and the introducing *ce/det* (or *c'est/det är*) serves as a forward pointer. Note that there is an affinity between backward-pointing (anaphoric) and forward-pointing constructions. An anaphoric construction (especially one with vague reference) can often be changed into a forward-pointing one by adding specifying information in the predication fol-

lowing *c'est/det är*. Consider Example (25) which is adapted from a dialogue with the learner C1:

- (25) (Interlocutor's question:
tyckte du att de va lätt att förstå henne?
'did you find it easy to understand her?')
(Alternative answers:)
- a. *de e lättare än att prata*
'it's easier than talking'
 - b. *de e lättare att förstå än att prata*
'it's easier to understand than to talk'

In (25a) *de* must be interpreted as an anaphoric pronoun, referring back to the interlocutor's *att förstå* 'to understand'. But if the pertinent information is supplied in the following context, as in (25b), the role of *de* switches to that of a forward-pointing dummy subject. Note that this latter construction is usually more explicit than the corresponding anaphoric construction; in fact the choice between a backward-pointing and a forward-pointing construction is an important way of regulating the degree of explicitness in cases like this and handling the balance between parsimony and clarity in formulation. The neutral pronoun itself loses or takes on a referring role accordingly. This contributes to the great flexibility of the collocation *c'est/det är*.

A forward-pointing *ce (c'est)* or *det (det är)* gives rise to a number of syntactically complex constructions such as Existential, Anticipating, Right Dislocation and Cleft. It is of interest to find out how often such 'non-canonical' or 'marked' sentence constructions with *c'est/det är* are actually used by the native speakers and the learners, and when the learners start making use of them. We noted above that the Swedish learners of French sometimes overgeneralize the use of *c'est* where standard French has *il est* or *il fait* (Type 12) or *il y a* (Type 13). We interpret this as the combined effect of the fact that (a) Swedish has *det är* in these cases, and (b) *c'est* is so widely applicable in other, comparable sentence constructions in French – a case that supports Andersen's 'Transfer to Somewhere' principle. According to this principle, transfer from L1 occurs to a significant degree

if and only if (1) natural acquisitional principles are consistent with the L1 structure or (2) there already exists within the L2 input the potential for (mis-)generalization from the input to produce the same form or structure. (Andersen 1983: 182; italics in original)

The third group in our taxonomy, C, comprises Cases (17)–(19) where a formulation break of some sort occurs after *c'est/det är*, followed by an iteration, a reformulation, or the abandonment of the current utterance. These types are interesting because they are the most obvious cases that suggest a role for

c'est/det är as a unit in formulation and shed light on the process of building utterances. We will discuss this matter in greater detail in a following section. For the present, we may just note that these phenomena are not limited to learner speech, but occur freely with native speakers as well. Thus, whether or not it is due to limited formulation proficiency that learners produce breaks in these cases, they here apply a method of building utterances which is also characteristic of native speakers' production.

As one of our referees pointed out, researchers have noted that natural learners of French L2 use the expression */se/* as a general predication in an early interlanguage system where they just have */se/* (= *c'est* 'it is or there is'), */jal/* (= *il y a* 'there is') and */jana/* (*il n'y en a pas* 'there is not') as verbal predications (see, e.g., Véronique 1994 and Trevisse 1986: 372: *ils /se/ très contents* = 'they are very pleased'). These researchers have proposed to analyse */se/* as an unanalysed sequence without the possibility to interpret it as containing a target-like cataphoric or anaphoric element. We have been aware of this analysis (see Bartning 1997c) and considered it for our data, but found that our learners of French do not seem to acquire such a system (for an exception, see however Example 22). Since all of them have participated in language courses, they are exposed to both oral and written input and readily perceive and learn morphology. This of course does not prevent learners from relying on a simple use of *c'est* at a beginner stage. Several beginners already use not only *c'est* but also *il est, ils sont*. They do not acquire a communicatively developed language without bound morphology. Furthermore they use the structure *X, c'est Y* already as beginners (see Table 4, below). In our study this sequence is analysed as a target-like left dislocation, in view of the very frequent use of the same structure in NS speech.

4.2. *The quantification of usage in the corpora*

4.2.1. *Quantitative analysis of c'est.* The distribution of the French constructions with *c'est* is shown in Tables 1–4 below, namely the use by the native speakers (Table 1), the university beginners, the secondary school students and the first term university students (Table 2). Table 3 presents the longitudinal development of *c'est*-structures by one beginner in five consecutive interviews. Table 4, finally, shows the longitudinal development in five consecutive interviews by four beginners during ten months.

As Table 1 shows, the *native speakers* use *all* the *c'est*-structures identified in our corpora by the taxonomy, except, of course, Structures 6, 8, 12 and 13 which do not exist in French. However, the Structures 12 and 13 are presented in all tables since they are used by Swedish learners of French.

The NS have an interesting distribution in their use: Structures 1:1 and 2 are the most frequent ones (16.3 % and 16.0 %, respectively) but other structures

Table 1. Quantitative distribution of *c'est* in the interviews of four native speakers

	Anne	Mél	P.-N.	Françoise	Total	%
1:1 C'est+NP	16	15	14	16	61	16.3
1:2 with relative	2	1	4	1	8	2.1
2. C'est+Adj	16	19	7	18	60	16.0
3. C'est+Adv/PP	7	5	8	8	28	7.5
4. C'est+Inf	–	–	1	–	1	0.3
5. C'est+Subcl	5	1	1	–	7	1.9
7. 'Tough' const.	–	–	–	2	2	0.5
9. Left disloc	12	9	10	19	50	13.4
10. Pseud.cleft	1	–	1	2	4	1.1
11. Term.Subcl.	1	1	–	–	2	0.5
*12. Impersonal	–	–	–	–	–	–
*13:1 Existential	–	–	–	–	–	–
13:2 with relative	–	–	–	–	–	–
14. Anticipating	16	8	4	7	35	9.4
15. Right disloc	2	2	–	2	6	1.6
16. Cleft	1	1	1	2	5	1.3
17. C'est+rep	5	5	15	18	43	11.5
18. C'est+refor	10	11	6	8	35	9.4
19. C'est+aban	9	5	8	5	27	7.2
Total	103	83	80	108	374	100 99.9
Text size	3808	3855	3162	2779	13604	
% of text	2.7	2.2	2.5	3.9	2.7	

Note: The structures marked with * do not exist in standard French, but examples of these cases do occur with learners in the subsequent tables.

such as Left dislocation (Structures 9, 13.4%), Anticipating (Structures 17, 9.4%), repeated *c'est* (Structures 17, 11.5%) and reformulated *c'est* (Structures 18, 9.4%) also attract many occurrences in NS speech. All these structures attract around 10% or more of the total numbers.

As Table 2 shows, the NNS do not use all the categories of *c'est*-structures but typically those which are most frequent in NS speech, namely Structure 1:1, 2 and 17. Beginners, e.g., only use seven out of the 15 French structures (plus one non-native structure, Structure 13:1) in their first interviews (Table 2). This tendency is confirmed by all the other 16 NNS in their first interviews: the most frequent structures are always the same, namely Structures 1:1, 2 and 17.

This state of affairs may be explained by several factors.

– Interlanguages often reflect target language frequency relations (Ellis 2002:

Table 2. *Synthesis of the use of c'est by the three learner groups in their first interviews in comparison with the NS*

	8 Beg	%Beg	8 Sec	%Sec	8 Uni	%Uni	Total NNS	%NNS	Total NS	%NS
1:1 C'est+NP	14	27.5	36	52.2	83	28.3	133	32.2	61	16.3
1:2 w. relative	—	—	—	—	3	1.0	3	0.7	8	2.1
2. C'est+Adj	20	39.2	14	20.1	65	22.1	99	24.0	60	16.0
3. C'est+PP	3	5.9	3	4.3	23	7.9	29	7.0	28	7.5
4. C'est+Inf	—	—	—	—	1	0.3	1	0.2	1	0.3
5. C'est+Subcl	—	—	—	—	7	2.4	7	1.7	7	1.9
7. 'Tough'	—	—	1	1.4	—	—	1	0.2	2	0.5
9. Left disloc	3	5.9	1	1.4	14	4.8	18	4.4	50	13.4
10. Pseud.cleft	—	—	—	—	—	—	—	—	4	1.1
11. Term.Subc	—	—	—	—	—	—	—	—	2	0.5
*12. Imperson	—	—	—	—	1	0.3	1	0.2	—	—
*13:1 Existent	2	4.0	2	2.9	4	1.4	8	2.0	—	—
13:2 with relat.	—	—	—	—	3	1.0	3	0.7	—	—
14. Anticipat.	—	—	2	2.9	10	3.4	12	2.9	35	9.4
15. Right dislo	—	—	—	—	4	1.5	4	1.0	6	1.6
16. Cleft	—	—	—	—	10	3.4	10	2.4	5	1.3
17. C'est+rep	4	7.8	7	10.1	40	13.7	51	12.3	43	11.5
18. C'est+refo	3	5.9	2	2.9	17	5.8	22	5.3	35	9.4
19. C'est+aba	2	4.0	1	1.4	8	2.7	11	2.7	27	7.2
Total	51	100	69	100	293	100	413	100	374	100
Text size	4011		4928		12959		22088		13604	99.9
% of text	1.3		1.4		2.3		1.9		2.7	

166, 179; Andersen and Shirai 1994), maybe in particular structures that are learnt as holistic chunks.

- The most frequent structures are also the syntactically most basic and simple ones which also are easier to acquire. Since they are more frequent, exposure is richer; the reason why they are frequent is that they are useful and easy to handle. The NNS early grasp the syntactically most basic/simple structures (Structure 1:1, 2, 3 in Tables 2–3). It is not until later that learners expand structures to more elaborate ones as 1:2 and 16 (Table 2, Uni. students).
- These ‘early’ structures, proportionally more frequent in NNS speech than in NS speech (NNS: Structure 1: 32 %, Structure 2: 24 %, NS: Structure 1 16.3 %, Structure 2: 16 %, Tables 1–2) substitute for other more complex and elaborated structures as already mentioned: Structure 1 *c’est NP* is sometimes a simplified structure to dodge choices of lexical verbs, subject pronoun and verb morphology, see example (24) above; Structure 2 *c’est Adj* is used to avoid choices of subject pronoun, gender and agreement with predicative adjectives as in examples (22) and (23) above; this tendency explains the high percentage of Structures 1:1 and 2 in NNS speech (cf. Tables 2–4). This tendency reflects the possibility of vague reference of *ce/det*.
- Structures 17 and 18 are also used early. They reflect *c’est* as a remarkable utterance starter since it offers so many possible ways of continuing the utterance. These structures are frequent both in NS and NNS productions (Table 2). It can also reflect a search for lexis and morphology as mentioned above (cf. Towell et al. 1996; Raupach 1984). Another interesting phenomenon reflected in the data and captured by the taxonomy is the following: the Structures 1: 2, 4, 5, 7, 10 and 11 are syntactically more complex than Structures 1: 1, 2 and 3 since they contain a subordinated or nominalised clause. Structures 14–16 also represent more complex syntactic structures. Our data show that the relatively complex structures are also used and learnt late in the development of *c’est*: in the first interviews of the beginners and the college students there are no instances of Structures 4, 5, 10, 11, 15, 16, just one example of 7 and two examples of 14 (Table 2).

Table 2 clearly shows that the university students are the only learners to use complex syntax as represented by Structures 1: 2 (with relatives), 4, 5 (with subclauses), 13: 2 (with relatives), 15 (Right dislocation) and 16 (Cleft). Two of the structures are not used by any learner, namely Structure 10 (Pseudo-cleft) and 11 (Terminal Subclause), only by the NS. The university students use Structure 10 in later interviews, as in example (10b) above, taken from Lena’s third interview (cf. Bartning 1997b). In the InterFra corpus it is only the university learners who are seen to expand *c’est NP* with a relative clause as in Structure 1:2 and 13:2 or who use Structure 16 (Cleft sentences; one exception). Evidence that syntactically more complex utterances (with subordination) are late developmental features in an IL is given elsewhere by Wolfe-

Table 3. Longitudinal distribution of *c'est* in five interviews during ten months by the beginner Robert

	Int 1	2	3	4	5	Total	%
1:1 C'est+NP	1	8	5	12	4	30	22.4
1:2 with relative	–	–	–	–	–	–	–
2. C'est+Adj	–	11	5	9	8	33	24.6
3. C'est+Adv/PP	–	6	8	2	5	21	15.7
4. C'est+Inf	–	1	–	–	–	1	0.7
5. C'est+Subcl	–	–	–	–	1	1	0.7
7. 'Tough'const	–	–	–	–	–	–	–
9. Left disloc	–	–	–	–	–	–	–
10. Pseud.cleft	–	–	–	–	–	–	–
11. Term.Subcl.	–	1	–	–	–	1	0.7
*12. Impersonal	–	–	–	–	–	–	–
*13:1 Existential	–	1	1	–	–	2	1.5
13:2 with relative	–	–	–	–	–	–	–
14. Anticipating	–	–	–	–	1	1	0.7
15. Right disloc	–	–	–	–	–	–	–
16. Cleft	–	–	–	1	–	1	0.7
17. C'est+rep	–	6	6	3	7	22	16.4
18. C'est+refor	–	4	2	6	4	16	12.0
19. C'est+aban	1	–	1	1	2	5	3.7
Total	2	38	28	34	32	134	100
							99.84
Text size	539	687	625	837	1016	3704	
% of text	0.4	5.5	4.5	4.1	3.1	3.6	

Quintero et al. 1998 for written IL and Bartning and Kirchmeyer (2003), for oral L2 production. Consequently, as the longitudinal profile of the beginner Robert's use of *c'est* during 10 months shows, in Table 3, there are just some isolated cases of Structure 4, 11 and 16.

The longitudinal tables show that there is an explosion of *c'est* after the first interview (Tables 3, 4). This indicates that *c'est* is hardly available for all beginners from the very start but soon after some more time and input: a learner has to have some more linguistic material available in order to use *c'est*. Tables 3–4 confirm the clear increase in Interview 2. Table 3 offers a longitudinal profile of one beginner, Robert. The profile confirms the discovery of *c'est*-structures in Interview 2 to the extent that Robert ventures even complex structures such as 4 and 11 and overuses *c'est* for *il y a* (Structure 13:1). In the last two interviews three of the complex structures emerge, Structures 5, 14 and 16.

Table 4. Synthesis of five consecutive interviews by four beginners (Carin, Emelie, Jan and Robert) during ten months

	Int 1	2	3	4	5	Total	%
1:1 C'est+NP	5	10	10	33	15	73	23.7
1:2 with relative	–	–	–	–	–	–	–
2. C'est+Adj	6	26	11	22	18	83	27.4
3. C'est+Adv/PP	3	9	9	6	7	34	11.0
4. C'est+Inf	–	2	1	2	–	5	1.6
5. C'est+Subcl	–	–	–	–	1	1	0.3
7. 'Tough' const	–	–	–	–	–	–	–
9. Left disloc	2	2	–	–	–	4	1.3
10. Pseud.cleft	–	–	–	–	–	–	–
11. Term.Subcl.	–	1	–	–	–	1	0.3
*12. Impersonal	–	–	1	–	2	3	1.0
*13:1 Existential	2	1	2	3	2	10	3.2
13:2 with relative	–	–	–	–	–	–	–
14. Anticipating	–	1	–	–	3	4	1.3
15. Right disloc	–	1	–	–	–	1	0.3
16. Cleft	–	–	–	1	–	1	0.3
17. C'est+rep	4	13	9	8	12	46	15.0
18. C'est+refor	3	5	3	8	8	27	8.8
19. C'est+aban	2	1	2	5	4	14	4.5
Total	27	73	48	88	72	307	100
Text size	1806	1797	1935	2520	2530	10588	
% of text	1.5	4.1	2.5	3.5	2.8	2.9	

4.2.2. *Quantitative analysis of det är.* The distribution of the Swedish constructions with *det är* is shown in Tables 5–8. As described earlier, the design of the Swedish and French corpora differs in some respects. To recapitulate, the Swedish data presented here were drawn from six native informants (NS, Table 5) and six learners (NNS, Table 6). The learners were all recorded in individual interviews during ten successive sessions ranging from near-zero to an advanced stage of proficiency, along a common time schedule, following the same content plan for the interviews. This forms the basis for tracing the longitudinal development for each learner as well as for the group. Thus, Table 7 gives an example of an individual longitudinal profile, and Table 8 shows the longitudinal group profile.

Two general observations can be made in studying the distribution of *det är* in the *native informants* (Table 5). First, the picture is very similar across the six informants, indicating a stable pattern of distribution for our NS. Second, there is a fairly great overall resemblance with the corresponding findings for

the French NS which were displayed in Table 1 above; this indicates a high degree of cross-linguistic similarity in target language usage between French and Swedish in the area that we are dealing with.

As is the case with the French NS, there are a few structure types which dominate in frequency. The structures 1:1 (*det är* + NP), 2 (*det är* + Adjective), 14 (Anticipating), and 17–19 (*det är* followed by some type of break) are among the most frequent types, like their French counterparts. Especially the two first types have wide applicability in that they form part of very basic sentence structures and can be used freely with a large number of different nouns and adjectives; therefore their high frequency should not be unexpected. It is worth noting that the Structures 17 and 18 and, to some extent, 19, i.e., cases with a syntactic break after *det är*, are amply attested with all the native informants. We take this as an indication that such breaks in the utterance formulation are a regular feature of spoken discourse and cannot be taken as a criterion of incomplete mastery of the language.

Certain structure types differ markedly in frequency between the Swedish and the French NS. Left dislocation (Structure 9) is much less used in Swedish (0.5 %) than in French (13.4 %). This reflects the fact that the characteristic French use of the *préambule*, although available, is less utilised by the Swedish NS (cf. also Conway 2005). On the other hand, the Cleft construction (Structure 16), is more prominent in Swedish than in French speech. This reflects the fact that Cleft is a very frequent and handy foregrounding mechanism in colloquial Swedish speech. An apparent difference in frequency occurs with Existential sentences (Structure 13:1–2). These occur frequently in Swedish. But since their counterparts in native French are constructed with *il y a* and not with *c'est*, the frequency of this type in French is not reflected in the present study.

The total frequencies and the percentages for the NS have also been included in the overview of the *learners' production* for comparison (Table 6). To a large extent, the totals for the NNS display a similar picture to that of the NS, the most prominent structures being 1:1, 2, 14, 17 and 18.

The fact that Structure 2, *det är* + Adjective, is even more prominent with the NNS (26.6 %, as compared to 17.4 % for the NS), is probably because the learners during the interviews tend to utter descriptive or evaluating judgments about things more often than the native speakers do. This is also manifested in an abundance of descriptive and evaluating adjectives in the learners' utterances, such as *dyr* 'expensive', *rolig* 'funny', *bra* 'good', *svår* 'difficult', *möjlig* 'possible', *viktig* 'important' and many others. Here Swedish and French learner usage reflects the same tendency.

Structure 9, *det är* after Left dislocation, is more common with the learners than with the native speakers (2.4 % vs. 0.5 %). On the other hand, the emphatic Cleft construction (Structure 16) is less used by the learners than by the native speakers (2.0 % vs. 8.2 %). This differs markedly from the French

Table 5. Quantitative distribution of *det är* in six native speakers of Swedish

Structures	NS						Total	%
	Z1	Z2	Z3	Z4	Z5	Z6		
1:1 <i>det är</i> + NP, no Rel	36	54	45	60	53	52	300	21.2
1:2 <i>det är</i> + NP, with Rel	4	18	5	4	13	5	49	3.5
2. <i>det är</i> + Adj	30	38	30	59	30	59	246	17.4
3. <i>det är</i> + Adv/PP	9	10	6	8	10	15	58	4.1
4. <i>det är</i> + Infinitive	-	3	1	1	-	-	5	0.4
5. <i>det är</i> + Subclause	4	6	11	6	1	1	29	2.1
6. <i>det är</i> w Stranded Prep	1	1	-	-	1	-	3	0.2
7. 'Tough' construction	6	2	2	2	5	2	19	1.3
8. Obj <i>det</i> Raised fr Subcl	-	1	-	1	-	-	2	0.1
9. <i>det är</i> w Left Disloc	-	2	2	1	1	1	7	0.5
10. Pseudo-Cleft	-	-	-	1	2	-	3	0.2
11. <i>det är</i> terminal in Subcl	6	3	2	8	-	1	20	1.4
12. Impersonal	3	2	2	2	4	1	14	1.0
13:1 Existential, no Rel	8	16	5	15	16	17	77	5.4
13:2 Existential, with Rel	5	19	17	9	2	4	56	4.0
14. Anticipating	18	22	22	22	18	12	114	8.1
15. Right Dislocation	1	-	1	1	1	3	7	0.5
16. Cleft	12	23	15	10	31	25	116	8.2
17. <i>det är</i> + repeated	9	18	13	21	28	5	94	6.7
18. <i>det är</i> + reformulation	25	26	10	44	28	15	148	10.5
19. <i>det är</i> + abandonment	10	8	11	3	7	8	47	3.3
Total <i>det är</i>	187	272	200	278	251	226	1414	100
Text size (word tokens)	12140	13562	10239	11532	15397	14367	77237	
% <i>det är</i> of text size	1.5	2.0	2.0	2.4	1.6	1.6	1.8	

Table 6. *Quantitative distribution of det är in six learners of Swedish*

Structures	NNS						Q1	Q2	NNS Total	NNS %	NS Total	NS %
	C1	C2	G2	G3	G3	Q2						
1:1 <i>det är</i> + NP, no Rel	33	109	52	64	48	48	354	21.3	300	21.2		
1:2 <i>det är</i> , with Rel	6	2	5	15	4	4	42	2.5	49	3.5		
2. <i>det är</i> + Adj	72	152	47	53	66	52	442	26.6	246	17.4		
3. <i>det är</i> + Adv/PP	15	17	16	8	12	2	70	4.2	58	4.1		
4. <i>det är</i> + Infinitive	-	7	2	-	4	-	13	0.8	5	0.4		
5. <i>det är</i> + Subclause	2	3	2	7	2	2	18	1.1	29	2.1		
6. <i>det är</i> w Stranded Prep	-	-	-	1	-	-	1	0.06	3	0.2		
7. 'Tough' construction	-	-	-	1	9	2	12	0.7	19	1.3		
8. Obj <i>det</i> Raised fr Subcl	-	-	-	-	-	-	-	-	2	0.1		
9. <i>det är</i> w Left Disloc	3	10	4	6	6	10	39	2.4	7	0.5		
10. Pseudo-Cleft	1	-	-	1	-	-	2	0.1	3	0.2		
11. <i>det är</i> terminal in Subcl	-	2	2	1	-	1	6	0.4	20	1.4		
12. Impersonal	10	2	5	2	9	4	32	1.9	14	1.0		
13:1 Existential, no Rel	10	15	1	4	18	5	53	3.2	77	5.4		
13:2 Existential, with Rel	-	2	5	4	4	1	16	1.0	56	4.0		
14. Anticipating	20	26	34	17	44	4	145	8.6	114	8.1		
15. Right dislocation	-	2	2	1	-	-	5	0.3	7	0.5		
16. Cleft	1	-	6	11	12	4	34	2.0	116	8.2		
17. <i>det är</i> + repeated	21	16	20	46	58	30	191	11.5	94	6.7		
18. <i>det är</i> + reformulation	19	42	16	18	36	7	138	8.3	148	10.5		
19. <i>det är</i> + abandonment	5	9	9	5	16	4	48	2.9	47	3.3		
Total	216	416	228	265	354	180	1661	100	1414	100		
Text size (word tokens)	11952	15077	13344	13678	12968	8363	75382		77237			
% <i>det är</i> of text size	1.8	2.8	1.7	1.9	2.7	2.2	2.2		1.8			

pattern. Like Left dislocation, Cleft is an optional foregrounding mechanism. Existential sentences with *det är* (Structure 13:1–2) are also less used by the learners. To some extent this can be accounted for by a certain overrepresentation at early stages of the near-synonym *det finns* ‘there is/there exists’.

A developmental picture appears if we look at the *quantitative profile over time* for each structure. An example of such a longitudinal overview for one individual learner, C1, is given in Table 7. Similar profiles have been obtained for the other five learners, but are left out here for reasons of space. They add up to a combined *developmental profile* for the group of six learners as shown in Table 8.

As Table 8 shows, the polyfunctionality of *det är* develops gradually in the learners’ production. At the very initial stage, in Session 1, *det är* occurs only in simple sentences followed by NP (Structure 1:1 and the existential Structure 13:1). But by Session 2 already, 12 of the 21 structure types (counting subtypes of 1 and 13) have appeared. By Session 6, 17 structure types are present, and by the end of the observation period, in Session 10, 20 types. Both anaphoric and non-referential use of *det är* are represented early on. The learners’ overall use of *det är* increases over time, as can be seen in the bottom row of the table. *Det är* becomes gradually even more frequent with the learners than with the native speakers. The amount of *det är* counted in the learners’ total text mass is 2.2 % vs. 1.8 % for the NS. The mean value for the last five sessions is as high as 2.4 %, i.e. one instance of *det är* per 42 text words on average.

Syntactic breaks after *det är* (Structures 17–19) occur regularly with all learners and, as Table 8 shows, they increase rather than decrease with higher proficiency. This supports our conclusion that these uses of *det är* are normal discourse phenomena; apparently the learners gradually reach stages of utterance complexity and fluency where these features become a regular part of discourse performance.

Whereas these various observations permit the general conclusion that *det är* soon becomes established as a favoured collocation in the speech of our learners, there is on the other hand also a striking difference between structures which tend to appear early in the learners’ production and structures which appear later, if at all. Thus, structures with anaphoric *det + är + NP*, Adjective, Adverb or PP (Structures 1:1, 2 and 3) occur earlier than, e.g., *det är + Infinitive* (Structure 4), *det är* with Stranded Preposition (Structure 6), Raised Object *det* (Structure 8, not occurring), Pseudo-Cleft (Structure 10), or terminal *det är* (Structure 11). Among the structures with non-referring or cataphoric *det*, Impersonal, Existential and Anticipating constructions (Structures 12, 13:1 and 14) appear earlier than Right Dislocation and Cleft (Structures 15 and 16). Such differences in time of development are especially obvious if we compare related structures, such as 1:1 vs. 1:2, 9 vs. 10, or 13:1 vs. 13:2, structures which differ essentially by the absence or presence of an embedded relative

Table 7. Individual longitudinal profile for the learner C1

Structures	Sessions										Total 1–10
	1	2	3	4	5	6	7	8	9	10	
1:1 <i>det är</i> + NP, no Rel	2	3	2	3		7	8		3	5	33
1:2 <i>det är</i> + NP, with Rel		1				2		1	1	1	6
2. <i>det är</i> + Adj		7		5	8	20	3	8	12	9	72
3. <i>det är</i> + Adv/PP			3			7		1	1	3	15
4. <i>det är</i> + Infinitive											–
5. <i>det är</i> + Subclause							1		1		2
6. <i>det är</i> w Stranded Prep											–
7. 'Tough' construction											–
8. Obj <i>det</i> Raised from Subclause											–
9. <i>det är</i> w Left Dislocation				1		1	1				3
10. Pseudo-Cleft										1	1
11. <i>det är</i> terminal in Subclause											–
12. Impersonal						1	1	2			10
13:1 Existential, no Rel				1	1	3		1		4	10
13:2 Existential, with Rel											–
14. Anticipating						2	2	6	3	7	20
15. Right dislocation											–
16. Cleft									1		1
17. <i>det är</i> + repeated		1	2	1		4	3	4	2	4	21
18. <i>det är</i> + reformulation			1	1	2	2	2	2	5	4	19
19. <i>det är</i> + abandonment		1				3	1				5
Total <i>det är</i>	2	13	14	12	11	52	22	25	29	38	218

Table 8. Longitudinal development, group frequencies for six learners

Structures	Sessions										Total 1–10
	1	2	3	4	5	6	7	8	9	10	
1:1 <i>det är</i> + NP, no Rel	18	30	10	26	14	50	58	26	37	85	354
1:2 <i>det är</i> + NP, with Rel		2		2	2	3	12	2	5	14	42
2. <i>det är</i> + Adj		21	7	34	47	73	58	38	75	89	442
3. <i>det är</i> + Adv/PP		5	3		1	14	11	7	15	14	70
4. <i>det är</i> + Infinitive						3	1		2	7	13
5. <i>det är</i> + Subclause		1		2		1	2	1	5	6	18
6. <i>det är</i> w Stranded Prep										1	1
7. 'Tough' construction				2	3	2	1	1		3	12
8. Obj <i>det</i> Raised from Subclause											–
9. <i>det är</i> w Left Dislocation		2		3	4	6	5	3	7	9	39
10. Pseudo-Cleft										2	2
11. <i>det är</i> terminal in Subclause						1		2		3	6
12. Impersonal		1	8	3		2	2	3	2	11	32
13:1 Existential, no Rel	8	1	5	5	2	9	1	2	2	18	53
13:2 Existential, with Rel							1		6	9	16
14. Anticipating		2	3	4	13	13	16	21	28	45	145
15. Right dislocation					2	1				2	5
16. Cleft				1		2	4	2	9	16	34
17. <i>det är</i> + repeated	3	12	7	25	20	19	22	17	22	44	191
18. <i>det är</i> + reformulation		1	6	5	13	10	24	12	31	36	138
19. <i>det är</i> + abandonment		1	6	2	5	7	11	3	3	10	48
Total <i>det är</i>	30	84	49	114	126	216	229	140	249	424	1661
Text size (word tokens)	2787	4239	5056	6505	5660	7735	8311	8241	9109	17744	75382
% <i>det är</i> of text size	1.0	2.0	1.0	1.8	2.2	2.8	2.8	1.7	2.7	2.4	2.2

clause. The degree of syntactic complexity evidently tends to have the same effect here as with the Swedish learners of French. Comparing early and late structures in Table 8, we can see a correlation between total frequency and time of appearance, the tendency being that those structure types which we have noted as less frequent in the speech of native speakers and learners, are seen to develop later in the learners' observed production. The difference between early and late structures is also evidenced by the number of learners who have started using the structure in question, as a comparison with Table 6 shows. Whereas those structures which we characterised as early are used by all the learners, the late ones are attested only with some.

In a comparison of the developmental data for the Swedish and French learners (Tables 2, 3, 4 for French, 6, 7 and 8 for Swedish), the distribution of early and late structures turns out to be similar to a fairly large extent. Common features are the early appearance of Structures 1:1, 2, 3, 9, 13:1, 17, 18 and 19, and the later emergence especially of Structures 4, 5, 10, 11, 13:2, 15 and 16. Structures 6 and 8, which are not possible in French, are "late" in Swedish, as evidenced by a single late instance of 6 and no occurrence of 8. Structures 12 and 13 are also cases where the two languages differ, and it is interesting to observe what happens here. These structures occur regularly in NS Swedish and appear early with the learners of Swedish; and despite the fact that NS French does not use *c'est* in these cases, constructions with *c'est* do occur with the Swedish learners of French. It may well be that the tendency for these structures to be carried over to learner French is favoured by their central role in Swedish L1 usage.

5. *C'est* and *det är* in utterance production

In the foregoing sections, we have dealt with the collocations *c'est* and *det är* from grammatical and quantitative-developmental points of view. In this section, we want to add the aspect of utterance production. In particular, we will consider the connections between the grammatical flexibility, the high frequency and the role of *c'est/det är* in the real-time process of forming utterances.

Kempen and Hoenkamp (1987) put forward the notion of *incremental processing* of utterances, a notion which also constitutes a central principle in Levelt's (1989) 'Speaking' model. Although utterances while being produced are assumed successively to run through the steps of conceptualising, formulating and articulating, the three sub-processes are thought to work in parallel, so that parts of the utterance are delivered continuously for the next processing component to start working with in the meantime. This serves the quick and fluent production of speech by allowing the utterance to get going while the speaker is planning how to complete it. The frequent occurrence of *c'est/det*

är as an introducer to various predications suggests that this collocation is a favoured standard sequence to make use of as a starter and an aid in the ongoing sentence-planning.

We believe that the developmental patterns which we demonstrated in Section 4 above can best be understood in the light of a *usage-based* conception of language, a model in which linguistic representations emerge as a result of experience and gradual *entrenchment as cognitive routines* (Langacker 1987, 1999; Kemmer and Barlow 1999). The fact that a wide range of useful syntactic patterns in French and Swedish share the common string *c'est* or *det är*, respectively, is crucial in our case. This creates a basis for *c'est* and *det är* to become established in the language-user's mind as syntactic *schemas* which are available and readily accessible as standard parts of many larger structures. According to a usage-based view, *frequency of use* plays a central role in this process, as the accumulated experience of repeated use of an item gradually leads to its entrenchment in the speaker's mind (Bybee and Hopper 2001a, b; Ellis 2002).

It is noteworthy that especially those structures with *c'est* and *det är* which are most frequent with the native speakers and the learners also tend to occur earliest in the learner corpora (see Tables 2, 3 and 4 for *c'est*; 7 and 8 for *det är*). We interpret this finding as the combined effect of functionality and frequency. The high frequency of some structures with our native speakers makes it reasonable to assume that the learners experience frequent use of these structures in their linguistic input from target language speakers. Frequent exposure should favour early entrenchment of the structure in the learner. On the other hand, the frequent use in the learner's own speech also favours the routinisation of these structures. The successive routinisation of a structure gradually makes it easier to process and hence more readily available to the user. Furthermore, if frequency of use is a function of communicative needs and usefulness, as we assumed above, then the polyfunctionality of *c'est/det är* should be an important basis for the routinisation of sentence patterns, the most useful structure types being the most driving in this respect.

Our data suggest that *c'est* and *det är* are entrenched as *sequences*. Ellis (1996) elaborates the idea "that much of language acquisition is in fact sequence learning and that abstract grammatical knowledge comes from analysis of sequence information" (Ellis 1996: 91). The learner's short-term memory recurrently analyses ordered verbal strings which are integrated in long-term memory from where they can be accessed in an automatic way. This involves *chunking*, "the development of permanent sets of associative connections in long-term memory" (Ellis 1996: 107).

Our findings support earlier observations in the literature that *c'est* and *det är* function as *formulaic sequences* in the ongoing planning and execution of utterances (Raupach 1984: 124; Axelsson 1988: 196 ff.). The notion of for-

mulaic sequence (variously called *formula*, *prefab*, etc.) is usually defined in the literature on the basis of how such items are thought to be accessed. In the words of Wray (1999: 214), formulaic sequences are “stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”. Formulas thus serve as planning units in speech production and occur as automatised chunks. They constitute holistic processes, interacting with creative processes in utterance encoding and decoding. As Wray and Perkins (2000: 11) put it,

the best deal in communicative language processing is achieved by the establishment of a suitable balance between creative and holistic processes. The advantage of the creative system is the freedom to produce or decode the unexpected. The advantage of the holistic system is economy of effort when dealing with the expected.

Unlike other common types of formulaic items, such as lexical phrases, and set phrases used in social interaction, *c'est* and *det är* are examples of *grammatical* formulas. These are characterised by Erman and Warren (2000:41) as “intralinguistic text-forming items rather than units with extralinguistic reference”. (For reviews of types and functions of formulas, see Raupach 1984; Weinert 1995; Erman and Warren 2000; Wray 1999, 2002; Wray and Perkins 2000; Forsberg 2006.)

One indication of formulaicity is the phonetic form which can reveal that the speaker retrieves and uses the item as a chunk. Formulaicity tends to favour a contracted or reduced form of pronunciation. *C'est* is always realised as a phonetically contracted unit, [se]. With *det är* there is a continuum of spoken variants: [de: e:, dee, de, d^e], where a more or less reduced form is what normally occurs. (The ASU corpus uses *de e* as a broad representation of this spoken-form continuum.) Coherence is obligatory, i.e., no pause is possible between *ce* and *est*, nor between *det* and *är*. Thus, repeated forms where *det* is followed by *det är* can take the shape *de de e*, or *de = de e* (where ‘=’ is an unfilled pause), or *de % de e*, (where ‘%’ is a pause-filling sound), but not **de de = e* or **de de % e*. Here *det är* (*de e*) is used as a chunk in a resumptive function after a break in the utterance. Similarly, the use of a resumptive *c'est* or *det är* is favoured in speech to form Left Dislocation and Pseudo-Cleft constructions, as in the examples under (9) and (10) above, rather than just continuing the sentence with *est* or *är*.

Particularly revealing evidence for *c'est* and *det är* as preliminary starters in the planning of an utterance are those cases where *c'est/det är* is followed by a syntactic break of some sort, i.e., our Structure types 17, 18 and 19 above. As we saw, there may be repetitions, reformulations or total abandonment of the utterance, corresponding to how well the speaker manages to develop a continuation which fits with the starter *c'est/det är*. Such items occur as typical

features in ongoing, coherent discourse, where the speaker is searching for formulations. Here is a Swedish example of how a speaker (the learner Q1) uses sentence-initial *det är* routinely in a piece of discourse while working out an answer to a question:

- (26) B: *ja. hur går de till å läsa arkitektur i i / på kth?*
 ‘well. how is it to study architecture at KTH?’
 I: *ja. de e de e väldigt spännande verkligen men också mycke jobbit. de e mycke mycke jobb. de e / man sitter där från nie till fem (B: MH) varje da. (B: MHM) och de e mycke intensivt kurs. kursen e mycke intensivt.*
 ‘well. **it’s it’s** very exciting really but also very tough. **there’s** much much work. **it’s** / you sit there from nine to five every day. and **it’s** [a] very intensive course. the course is very intensive.’
 (Q1-10)

Reformulations of various kinds (Structure type 18) are of special interest, because they show clearly how a more developed utterance is successively built up with the help of an unspecific starter *c’est/det är*, and thus offer insights into the gradual process of conceptualisation and formulation.

We will illustrate this with examples from NS and NNS speech of cases where *c’est/det är* is followed by a syntactic break and a reformulation of some kind. In Swedish, the starter may also contain an unstressed modal adverb (*det är ju, det är väl, det är liksom*, etc) within the same prosodic unit.

The reformulation can be a minor or a more radical one. For example, it may consist of just a shift of tense, as in (27a) or aspect, as in (27b). Even such a minimal change illustrates the formulaic use of the starter *c’est* or *det är*.

- (27) a. *c’est c’était à la fois littéraire et historique.*
 ‘**it’s it was** at the same time literary and historical’ (NS: Anne Int)
 b. *vi snackar tillsammans skojar tillsammans. (B: JA) de e / de blir kul när vi jobbar tillsammans.*
 ‘we chat together joke together. (B: YES) **it’s / it gets** fun when we work together’ (G3-10)

Many of the reformulations reveal the gradual build-up of an argument structure for the sentence. The speaker sets out with the provisional *c’est/det är*, and then elaborates the sentence with a new verb and/or a specific subject noun or pronoun. Some examples are shown in (28).

- (28) a. *en même temps c’est # ça leur permet peut-être de de de se donner de la liberté de cette façon.*
 ‘at the same time **it’s # this permits** them to to ot give themselves freedom in this way’ (NS: Anne Int)

- b. *j'aime pas le télé parce que c'est # le télé on ne / on est toute seule avec le télé.*
 'I do not like television because **it's # television you not / you are totally alone with television**' (Uni: Malin Int1)
- c. *jo men de e / ja e väldigt intresserad av ganska mycke.*
 'yes but **it's / I'm** very interested in quite a lot of things' (Z4-1)
- d. I: *ja tåget.*
 'I yes the train'
 E: *ja.*
 'yes.'
 I: *men lastbile+ de e också / man kan kombinera tåg å lastbil*
 'but the lorry **it's also / you can combine** train and lorry'
 (C2-8)

Other common types of modifying the *det är* clause are topicalisation of a non-subject (29a–b) and embedding the clause by inserting a verb of opinion (29c–d):

- (29) a. *ja de e / ibland = e de = orättvist tror ja.*
 'yes **it's / sometimes = it's = unfair I think.**' (C2-9)
- b. *ja tycker att de e / på nå sätt / ur moralisk synvinkel så e de väl ganska så förkastligt*
 'I think that **it's / in a way / from a moral point of view I suppose it's** rather objectionable' (Z1-3)
- c. *ja precis. de e / ja anser att de e så.*
 'yes precisely. **it's / I think that it's** so.' (G3-8)
- d. *ah oui à l'université oui c'est bien / c'est # je crois que c'est onze livres / en français.*
 'ah yes at the university yes **it's OK / it's # I think that it's** eleven books in French' (Uni: EvaInt1)

An interesting case of French *c'est* as a starter is the switch from *c'est* to *il y a*, especially in learner discourse, but quite regular also in native speech. It is not only *il y a* but also other impersonal constructions with *il* as formal subject that follow after the *c'est* starter even in native speech:

- (30) a. *quand j'habitais à Val de Loire c'est # il y a beau- / il y avait beaucoup de: de vin / qui sont bien. (I:mm)*
 'when I lived in the Loire Valley **it is # there was much / there was** much wine which was good.' (Uni: Eva Int1)
- b. *l'hiver c'est # il fait quinze degrés donc.*
 'in winter **it's # it's** fifteen degrees thus' (NS: Françoise Int)

- c. *c'est c'est # il faut faire attention quand on est en voiture.*
'it's it's # you have to be careful when you are in a car' (NS:
Françoise Int)

As we can see from the examples in (27), (28), (29) and (30), *c'est/det är* tends to be used as a general point of departure in working out a message. In some cases, such as (27a–b), the reformulation limits itself to a relatively minor change. But in other instances, the sentence is restarted in a more radical way, suggesting that essential parts of the conceptual information for the remaining utterance were not yet available at the point of the break. Sometimes a fundamentally different plan is devised for the continued utterance, and the schema *c'est/det är* is discarded. In many cases, however, there seems to be a tendency to keep the schema *c'est/det är*, at least in part, when working out the rest of the utterance.

6. Conclusion

In this study of functionality, formulation and frequency, we have chosen to focus, in each of two languages, on a collocation which is extremely polyfunctional as a component in different grammatical constructions, plays a significant role as a formula in the planning and execution of utterances, and has a characteristic distribution of frequency of use. We have tried to show how these properties are connected to each other, and how they together have a bearing on the development of learners' interlanguages. In a sense we have taken an opposite point of view compared to what has been usual in second-language studies, when we have chosen not to look at more difficult or advanced phenomena which limit the learner's L2 competence, but rather to concentrate on a very central item in language usage, one that occurs all the time and is readily acquired by learners. *C'est* and *det är* are obviously very useful standard elements in discourse and in utterance production. It seems that speakers need the access to such standard elements for their formulation work. The acquisition and further refinement of such central linguistic means is an important part of the development of second language competence.

Stockholm University
<Inge.Bartning@fraitu.su.se>
<ham@ling.su.se>

Appendix: Transcription conventions in the examples

When reading the examples, please note the following differences between the *InterFra* and the *ASU* practice.

The French InterFra corpus

Source references with informant and recording session are given in parenthesis after each example; e.g.: speaker identification (group and pseudonym: Beg(inner): Carin), task and chronological number of the task: Int(erview) 1, BD 1 = ‘bande dessinée’= cartoon.

The letters I and E in the left column are speaker identifications used where the text extract contains turns from more than one speaker. Here I refers invariably to the interviewer and E to the (native or non-native) informant.

Special symbols used:

/	Short empty pause
//	Longer empty pause
#	Reformulation marker
?	Completed question
.	Completed sentence other than question
< >	Around a code switch
(SIGH)	Non verbal sounds
+	Beginning of simultaneous speech
SIM	End of simultaneous speech
()	Around interlocutor’s backchannelling

The Swedish ASU corpus

Source references with informant and recording session are given in parenthesis after each example; e.g., Z4-1 = informant Z4, session no. 1; C2-6 = informant C2, session no. 6. The letter code Z refers to native Swedish informants; the letter codes C, G and Q refer to Chinese, Greek and Portuguese learners, respectively.

The letters I, B and E in the left column are speaker identifications used where the text extract contains turns from more than one speaker. Here I refers invariably to the (native or non-native) informant in question, and B and E are the interviewers. If no such speaker identification is given, the text extract is from the informant.

Special symbols used:

=	Empty pause
%	Pausefiller (replaces letter notations such as “eh”, “ah”, “ööh”)
-	Interrupted word
+	Morphologically unclear ending
/	Interruption, reformulation; non-completed sentence
?	completed question sentence
.	completed sentence other than question
< >	Around a code-switch
()	Around interlocutor’s backchannelling

References

- Allwood, Jens (1999). *Talspråksfrekvenser: frekvenser för ord och kollokationer i svenskt tal- och skriftspråk*. (Gothenburg Papers in Theoretical Linguistics S 21.) Gothenburg university, Department of linguistics.
- Andersen, Roger W. (1983). Transfer to somewhere. In *Language Transfer in Language Learning*, Susan Gass and Larry Selinker (eds.), 177–201. Rowley MA: Newbury House.
- Andersen, Roger W. and Yasuhiro Shirai (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16: 133–156.
- Axelsson, Monica (1988). Helfraser och ramar – ett viktigt inslag på andraspråksinlärares väg mot ett kreativt regelsystem. In *Första symposiet om svenska som andraspråk. Vol. 1: Föredrag om språk, språkinläring och interaktion*, Kenneth Hyltenstam and Inger Lindberg (eds.), 191–202. Stockholm University, Centre for Research on Bilingualism.
- Bartning, Inge (1997a). L'apprenant dit avancé et son acquisition d'une langue étrangère. Tour d'horizon et esquisse d'une caractérisation de la variété avancée. *AILE* 9: 9–50.
- (1997b). *C'est* in native and non-native spoken French. In *Norm, Variation and Change in Language. Proceedings of the Centenary Meeting of the Nyfilologiska Sällskapet. Studier i modern språkvetenskap*, vol. 11: 13–47. Stockholm: AWE International.
- (1997c). Structuration des énoncés et stratégies référentielles à l'aide de la prédication *c'est X* chez des apprenants avancés et des locuteurs natifs. *Travaux de linguistique* 34: 65–90.
- (2002). La variété avancée et sa place dans les phases acquisitionnelles d'une langue étrangère. In *Mélanges publiés en hommage à Gunnel Engwall*, Inge Bartning, Johan Falk, Lars Fant, Mats Forsgren, Ritva Maria Jacobsson and Jane Nystedt (eds.). Acta Universitatis Stockholmiensis. Romanica Stockholmiensis 20: 1–12. Stockholm: Almqvist and Wiksell International.
- (2006). Une formule bien utile : le cas de *c'est*. In *Aux carrefours du sens. Hommages offerts à Georges Kleiber pour son 60^e anniversaire*, Martin Riegel, Catherine Schnedecker, Peter Swiggers and Irène Tamba (eds.), 175–190. Louvain: Peeters.
- Bartning, Inge and Nathalie Kirchmeyer (2003). Le développement de la compétence textuelle à travers les stades acquisitionnels en français L2. *AILE* 19: 9–39.
- Bartning, Inge and Suzanne Schlyter (2004). Itinéraire acquisitionnel et stades de développement en français L2. *Journal of French Language Studies* 14: 281–299.
- Beckman, Natanael (1934). Västeuropeisk syntax. Några nybildningar i nordiska och andra västeuropeiska språk. In *Göteborgs högskolas årsskrift*, Vol. 40, No. 1934/4: 1–44. Göteborg.
- Blanche-Benveniste, Claire (1990). *Le français parlé. Études grammaticales*. Paris: CNRS.
- Brown, Gillian and George Yule (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Bybee, Joan and Paul Hopper (2001). Introduction to frequency and the emergence of linguistic structure. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee and Paul Hopper (eds.), 1–24. Amsterdam/Philadelphia: John Benjamins.
- Bybee, Joan and Paul Hopper (eds.) (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: John Benjamins.
- Conway, Åsa (2005). Le paragraphe oral en français L1, en suédois L1 et en français L2. Étude syntaxique, prosodique et discursive. *Études romanes de Lund* 73. PhD thesis. Lund: KFS AB.
- Ekerot, Lars-Johan (1979). Syntax och informationsstruktur. In *Svenska i invandarperspektiv*, Kenneth Hyltenstam (ed.), 79–108. Lund: Liber Läromedel.
- Ellis, Nick (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18: 91–126.
- (2002). Frequency effects in language processing. *Studies in Second Language Acquisition* 24: 143–188.
- Erman, Britt and Beatrice Warren (2000). The idiom principle and the open choice principle. *Text* 20: 29–62.

- Forsberg, Fanny (2006). Le langage préfabriqué en français parlé L2. Étude acquisitionnelle et comparative. Ph.D. Thesis. Stockholm University, Department of French, Italian and Classical Languages.
- Fraurud, Kari (1992). Processing noun phrases in natural discourse. PhD thesis. Stockholm University, Department of Linguistics.
- (2000). Demonstrativer i svensk sakprosa. In *Denna – den här – den där. Om demonstrativer i tvärspråklig belysning. En minnesskrift till Elsie Wijk-Andersson*. Special issue of *ASLA Information* 26 (2): 5–28. Uppsala: ASLA [Association Suédoise de Linguistique Appliquée].
- Fraurud, Kari and Christina Hellman (1999). Abstract objects: Transitional discourse referents. Paper presented at the 6th International Cognitive Linguistics Conference, 10–16 July 1999, Stockholm. (Quoted from handout.)
- Gadet, Françoise (1989). *Le français ordinaire*. Paris: Colin (2nd ed. 1997)
- Hammarberg, Björn (1999). Manual of the ASU-corpus, a longitudinal corpus of adult learner Swedish with a corresponding part from native Swedes. Department of Linguistics, Stockholm University.
- (2000). A polyfunctional word in native usage and L2 acquisition: the Swedish pronoun ‘det’. In *Kontraster i språk*, Johan Falk, Gunnar Magnusson, Gunnel Melchers and Barbro Nilsson (eds.), 103–129. Stockholm: AWE International.
- Hammarberg, Björn and Åke Viberg (1977). The place-holder constraint, language typology, and the teaching of Swedish to immigrants. *Studia Linguistica* 21: 106–163.
- Hancock, Victorine (2000). Quelques connecteurs et modalisateurs dans le français parlé d’apprenants avancés. Étude comparative entre suédophones et locuteurs natifs. Cahiers de la recherche 16. Stockholm University. Ph.D. thesis.
- Holmes, Philip and Ian Hinchliffe (2003). *Swedish: A Comprehensive Grammar*. 2nd ed. New York: Routledge
- Kemmer, Suzanne and Michael Barlow (1999). Introduction: A usage-based conception of language. In *Usage-Based Models of Language*, Michael Barlow and Suzanne Kemmer (eds.), vii–xxvii. Stanford: CSLI Publications, Center for the Study of Language and Information.
- Kempen, Gerard and Edward Hoenkamp (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science* 11: 201–258.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- (1999). A dynamic usage-based model. In *Usage-Based Models of Language*, Michael Barlow and Suzanne Kemmer (eds.), 1–63. Stanford: CSLI Publications, Center for the Study of Language and Information.
- Levelt, Willem J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge MA: The MIT Press.
- Morel, Mary-Annick and Laurent Danon-Boileau (1998). *La grammaire de l’intonation. L’exemple du français*. Paris: Ophrys.
- Olsson, Hugo (1986). *La concurrence entre il, ce, cela (ça) comme sujet d’expressions impersonnelles en français contemporain*. Acta Universitatis Umensis 71. Stockholm: AWE International.
- Pedersen, John, Ebbe Spang-Hanssen and Carl Vikner (1982). *Fransk universitetsgrammatik*. Stockholm: Esselte Studium.
- Raupach, Manfred (1984). Formulae in second language speech production. In *Second Language Productions*, Hans W. Dechert, Dorothea Möhle and Manfred Raupach (eds.), 114–147. Tübingen: Gunter Narr.
- Riegel, Martin, Jean-Christoh Pellat and René Rioul (1994). *Grammaire méthodique du français*, Paris: PUF.
- Sundman, Marketta (1980). *Existentialkonstruktionen i svenskan*. Publications of the Research Institute of the Åbo Akademi Foundation, 57. Åbo.
- (1987). *Subjektval och diates i svenskan*. Åbo: Åbo Academy Press.

- Teleman, Ulf, Staffan Hellberg and Erik Andersson (1999). *Svenska Akademiens Grammatik*. Vol. 1–4. Stockholm: Norstedts.
- Towell, Richard, Roger Hawkins and Nives Bazergui (1996). The development of fluency in advanced learners of French. *Applied Linguistics* 17 (1): 84–119.
- Trévisé, Anne (1986). L'émergence et la spécification des marqueurs de prédication en langue 2. In *Acquisition d'une langue étrangère : perspectives et recherches*. Alain Giacomo and Daniel Véronique (eds.), 365–381, Aix-en-Provence: Université de Provence.
- Weinert, Regina (1995). [The role of formulaic language in second language acquisition: A review](#). *Applied Linguistics* 16: 180–205.
- Véronique, Daniel (1994). *Il y a et c'est* dans l'interlangue d'apprenants de français. *Cahiers du français contemporain* 1: 55–75. Paris: Didier Érudition.
- Wolfe-Quintero, Kate, Shunji Inagaki and Hae-Young Kim (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity*. Honolulu: University of Hawai'i Press.
- Wray, Alison (1999). Formulaic language in learners and native speakers. *Language Teaching* 32: 213–231.
- Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison and Michael R. Perkins (2000). The functions of formulaic language: An integrated model. *Language and Communication* 20: 1–28.