

Subject Drop in Swiss French Text Messages

Elisabeth Stark & Aurélia Robert-Tissot, Zurich*

Printed versions available under: <https://benjamins.com/catalog/lv>

Abstract: In this paper, we present evidence in favour of a syntactic approach to subject drop in Swiss French text messages. Subject drop in our corpus follows patterns found in various so-called "written abbreviated registers" such as diaries, notes etc.: it occurs at the beginning of main sentences and after preposed adjuncts. Based on a corpus of 1100 text messages, collected in 2009/10 (www.sms4science.ch), we test predictions put forward by two approaches to argument drop in abbreviated registers, i.e. the "Avoid Weak Start" hypothesis by Weir (2012a) and the "Truncated CP hypothesis" by Haegeman (2013). While for our data the first approach cannot be excluded, our results more strongly support the syntactic one, despite the fact that some data, especially preposed strong subject *moi* without clitic resumption, challenge existing analyses. These data suggest that dropped referential subjects can be analysed as instances of familiar topic drop.

Key words: Text messages, (Swiss) French, subject drop, abbreviated registers, English, diary writing

1 Introduction – text messages as an abbreviated register

Argument and article drop are typical register markers in so-called written *abbreviated registers* such as headlines, recipes, or diaries; see the contributions by Weir, and in particular Haegeman, in this volume. They also appear quite regularly in text messages; for comments on ellipsis in computer-mediated communication in general see: Androutsopoulos & Schmidt 2002, Panckhurst 2009, Fairon et al. 2006. This is illustrated by example (1), where the utterance- and sentence-initial subject, *Je* 'I', is dropped in a French text message from our

* This research was financed by Grant Nr. CRSII1_136230 from the Swiss National Science Foundation. We also gratefully acknowledge the highly valuable comments on our talk presented at the 19th Congress of Linguists in Geneva, July 2013, and very stimulating discussion with Liliane Haegeman and Andrew Weir. We thank Mathieu Avanzi for prosodic advice. All remaining errors are, of course, ours.

corpus of Swiss text messages (*sms4science.ch*). (Unless otherwise marked, all example sentences are in French)

- (1) Ø Suis au boulot... dommage... tes sur ke vendredi
(Je)
(I) am at.the work... pity... you.NOM=are sure that Friday
- Ø Ø yapud' place?
(il n') (y a plus de)
(it not) there=has.no longer.of place

'I am at work. What a pity. Are you sure that there is no room (available) on Friday?'

Note that in this example, every constituent except the missing subject behaves as in colloquial French (if we ignore typical orthographic deviations from the norm, cf. Anis 2007 for French text messages). It is this type of example that we are interested in, and not completely reduced messages like the following, written in so-called "telegraphic style" where every functional element can potentially be omitted (cf. Hageman 2013:93, footnote 4):

- (2) Magnif l afrique super dépaysement et dur dur le retour!
wonderful the Africa great change-of-environment and hard hard the return
- 'Wonderful, Africa, excellent change, and so hard the return!'

The term *register* is used here to mean "situationally defined varieties" (Biber 1995:1), in contrast to speaker-dependent *dialects* (regional or social). Important extra-linguistic factors defining registers and triggering variation between different registers are, following Biber (1995), communicative intention, *interaction* (which will be of central interest for our empirical findings), conditions of use, and the type of relation between interlocutors. As early as 1982 so-called *abbreviated or "simplified" registers* (e.g. lecture notes or motherese) were described by Charles Ferguson from a theoretical perspective. Among the reasons for simplification, Ferguson mentioned facilitating acquisition, reducing cognitive load, economy of time and effort, etc. (1982: 59). Simplified registers are relevant for linguistic theory, as

whatever we find out about human language also has to be applicable to these registers, and, vice versa, they might help considerably in understanding general cognitive processes of language production and processing. Simplification can manifest itself as the “reduction of surface-phonetic material” (“length reduction”) or of “*grammatical* complexity” (cf. Janda 1985:452, following Hymes 1971), the latter recalling Rizzi’s “Avoid Structure” principle (Rizzi 1997:314).

The simplification or drop processes observed (argument and article drop) appear most strikingly in written or more precisely *graphic* abbreviated registers (diaries, notes, recipes, headlines, cf. Weir 2012a:105), i.e., certain cases of reduction seem to be bound to the medial (phonic or graphic transmission) character of abbreviated registers (see the denomination “written subject omission” (WSO) in Haegeman’s contribution to this volume). The graphic character is also evident for most of the new communication forms usually subsumed under the heading of “computer-mediated communication” (cf. Walther 1996), among which text messages are a prominent example. Compared to the other abbreviated graphic registers mentioned above (recipes, diaries and so on), text messages stand out, however, by *two important characteristics* that make them special: first, they are close to phonetics (cf. the “three maxims of texting”: “brevity and speed; paralinguistic restitution; *phonological approximation*”, Thurlow & Poff 2013:176; see also Anis 2007: “phonetic spelling”; but see Weir 2009:25ff., against assuming purely pragmatic economy for this). Second, they are highly *interactional* and dialogical in nature (but still a type of asynchronous communication as opposed to face-to-face-interaction, see Herring 2007).

Text messages are thus subject to both economy constraints *and* expressivity constraints (cf. Bieswanger 2013). While this may go some way towards explaining certain linguistic features that they exhibit, such as argument or article drop, it is not sufficient to grasp their detailed syntactic properties.

In what follows, we will concentrate on one specific phenomenon of argument drop, namely subject drop, in a corpus of French text messages, and compare it to what has been found in the literature on syntactic variation about subject drop in other abbreviated registers. We then discuss two prominent explanatory approaches to register-specific subject drop in order to see how far we can get when applying them to our data.

The first, proposed by Haegeman, is, to our knowledge, the only one based on authentic corpus data of abbreviated registers (including a quantitative analysis, see Haegeman & Ihsane 1999, 2001), proposing an elaborate syntactic analysis of them.¹ Teddiman (2011), Teddiman and Newman (2007), as well as Oh (2005, 2006) and Nariyama (2004) also conducted corpora studies into subject drop, mainly in spoken English, but their approach is pragmatic (i.e. interactional), mainly claiming that the recoverability of the subject referent is a necessary condition for subject ellipsis. However, this cannot be a satisfactory explanation: first, not all recoverable subjects are dropped, and second, as shown in example (3), subjects are dropped even when the resulting expression is ambiguous.

- (3) Ø *Voulait* *juste savoir* *si t'étais* *dans le train!*
 (Je/Il/Elle)
 (I/He/She) wanted.3.SG just know.INF if you=were.2.SG in the train
- Tant pis...*
never mind

‘(I/He/She) just wanted to know if you were on the train! Never mind...’

Morphologically, the initial verb *voulait* is marked for a 3rd person singular subject (-t). The phonic realization of this form is homophonous with the forms for the first and second person singular ([vule], spelt <voulai-s>), which could also have been the intended meaning. The author of this message does not seem to worry about this ambiguity and drops the subject despite the potential for misunderstanding. For these reasons, in this article, as in Haegeman’s work, based on English and French data –both European SVO languages with obligatory

subjects–, we focus on the *syntactic* regularities underlying subject drop in our Swiss French text messages. We will discuss this approach to the phenomenon rather than available studies on subject drop in languages which generally allow argument drop, such as Russian (partial *pro*-drop, cf. Zdorenko 2010) or Japanese (*topic*-drop, cf. Nariyama 2013).

The second approach we will discuss is the prosodic approach to argument drop (subject drop) proposed by Weir (2012a) for English, in view of the typological similarities mentioned above with our data and the fact that text messages are often considered to represent spontaneous speech more closely than more near-standard forms of writing (cf. Anis 2007, Stark 2015 for a summary). We will conclude with two problematic findings in our corpus for any approach, which we will leave for future research.

The paper is structured as follows: the next section, section 2, will introduce two prominent analyses of subject drop in abbreviated written registers, one prosodic, post-syntactic, the other one syntactic in nature. Section 3 briefly presents the data our analysis is based on, namely a sub-corpus of the Swiss reference corpus of text messages, *sms4science.ch*, while section 4 will summarize the main empirical findings concerning the distributional properties of subject drop in our corpus of French text messages, which parallel mostly, but not completely, those of Haegeman (2013 and this volume) for diary subject omission. Section 5 discusses these findings against the two approaches to subject omission presented in section 2 and focuses on one structure, stressed subject pronouns without a coreferential subject clitic (1st person). This structure is very rare in spoken French but recurrent in the text messages corpus and incompatible with the ‘reduced CP analysis’ put forward by Liliane Haegeman. Section 6 summarizes the main points and mentions some issues for future research.

2 Register-specific subject drop - some explanatory proposals and questions to ask

2.1 Postsyntactic Left-edge deletion in spoken English: “Avoid weak start” (Weir 2012a)

One explanation for subject drop in colloquial spoken English and also in abbreviated written English registers such as diaries in the vein of “reduction of surface material” (length reduction, see introductory remarks above) is the prosodic explanation presented in Weir (2012a), based on work by Napoli (1982). Subject drop (and other phenomena) is explained as a general postsyntactic preference to start prosodic phrases by stressed rather than weak material, with the result that weak syllables before the main stress of a prosodic unit tend to be dropped. This can be illustrated by the French example (4) from our corpus, as well as by its English word-by-word translation of the second sentence:

- (4) *Salut, comment vas tu ? Ø Ø Bien terminé la soirée ?*
 hello how go.PRS.2.SG you have.PRS.2.SG=you well finished the evening
 (As-tu)
 ‘Hi, how are you? (Have you) Finished the evening?’

In French, main stress (primary stress) falls on the last metrical syllable of prosodic words, i.e. on the last masculine syllable (i.e. a syllable whose nucleus is not a schwa, see Dell 1984) of lexical or content words (adverbs, nouns, adjectives, and verbs, and also some pronouns (*tout, chacun*), tonic pronouns (*moi, toi, eux*), etc., as well as some polysyllabic prepositions (*pendant, depuis*, which can be uttered in isolation). Primary stress delimits the boundary of minimal prosodic groups, which are called *Phonological Phrases* (Nespor & Vogel 1986; Delais-Roussarie 1996) or *Accentual Phrases* (Jun & Fougeron 2000). All other syllables (non-final lexical words or function word syllables) cannot receive primary stress, and are considered weak syllables. Thus, both the clitic subject *tu* (‘you’) and the auxiliary *as* (‘have’) in the second sentence of example (4) are prosodically weak, and their presence in front of the first stressable element, *bien* (‘well’), would violate Weir’s constraint as well as prosodic preferences observed independently for French (cf. Jun & Fougeron 2000), i.e. avoiding more than one unaccented syllable before the first stress of a prosodic unit. The next example (5) of

French subject drop in diaries represents a counterexample to Weir (2012a), who formulates, however, his proposal only for colloquial English. Example (5) is still completely congruent with Jun and Fougeron's (2000) predictions for French:

- (5) \emptyset *Me dit que l'architecte Perret est désireux de passer*
(Il/Elle)
(He/She) me.DAT says that the architect Perret is desirous of spend.INF

un moment avec moi

a moment with me

'He/She tells me that the architect Perret would love to spend a moment with me.'

(Haegeman 1997:234; Léautaud, *Journal particulier* 1933:44)

Of course, we have to be very careful here and cannot transfer prosodic explanations from phonic to graphic data, bearing in mind that especially in languages with highly conventionalized and standardized writing systems such as English or French, we cannot in general assume any *direct mapping relation between sound and spelling* (cf. for further discussion Béguelin 2012, Stark 2015). Additionally, some drop phenomena that are ungrammatical in spoken varieties such as the impossibility of subject drop with cliticizable auxiliary forms like *is* ('s) or *have* ('ve) in English or with preposed adjuncts in English or French are attested in graphic (written) abbreviated registers (see Weir 2012a and Haegeman this volume). However, registers like text messages with quite 'standard-free' forms of writing do show some 'phonetic spelling', as has been attested repeatedly in the literature (see, among many others, Anis 2007 for French).

The first question we will attempt to answer for our corpus, which is graphic in nature but very close to phonetics (see introduction), is thus based on the postsyntactic, prosodic assumption, following Weir (2012a) for colloquial English and English diaries and Jun and Fougeron (2000) for French, that French prefers at most one unaccented element in front of the first stress of a prosodic unit. Do we find empirical evidence for the systematic drop of unstressed first elements in a colloquial register like text messages?

2.2 *Syntax; Reduction of grammatical complexity: “reduced CP” (cf. Haegeman 1997, 1999, 2013, Weir 2012b)*

The second explanation for subject drop in abbreviated registers is, unlike the preceding one, truly syntactic in nature and based on specific assumptions concerning the left periphery of sentences. Haegeman (2013 and this volume) assumes, following Cardinaletti 2004 and Rizzi and Shlonsky (2007), a functional projection SubjP above TP and TP-adjoined adverbials, where canonical sentence subjects are located. This projection is occasionally allowed to be the highest phase head of a sentence, i.e., the root phase (cf. Chomsky 2001), as subject drop is a phenomenon of root clauses only. Accordingly, material in its specifier (= subjects in diaries etc.) cannot be spelt out. In this view, the articulated CP (following Rizzi 1997) is considerably reduced, possibly because in certain registers such as diaries or text messages, it is often less necessary to link sentences to the discourse context via e.g. topic-shift (= the function of CP), since the topic is given most of the time (speaker/writer in diaries, author, addressee or a third well-known entity in text messages etc.). Subject drop is compatible, among other things, with preposed adjuncts, but not, however, with subject-auxiliary inversion, *wh*-elements, complementizers or fronted arguments, located in functional projections above TP and also SubjP, which makes this analysis highly plausible. A related idea can be found in Weir (2012b), who assumes an inactive CP with respect to agreement, so that there will be no moved material in it (and accordingly above TP), only adjuncts. Neither approach excludes the possibility that sometimes and even most of the time, CP is projected and active for agreement also in abbreviated registers (cf., e.g., Haegeman 2013:105).

The second question we will try to answer is thus the following one: do we find parallel syntactic properties of subject drop in French text messages to the cases of diary (written) subject omission described in Haegeman (2013 and this volume), i.e. a neat preference or even categorical distribution as to the root-embedded asymmetry,

incompatibility with inverted auxiliaries, preposed *wh*-elements and arguments, but occurrences with preposed adjuncts/adverbials?

3 *Data*

The data we present here were collected in Switzerland from September 2009 to January 2010 with a supplementary wave for more Italian and Romansh data between April 2011 and July 2011 (cf. www.sms4science.ch). Citizens from all over Switzerland were invited via a huge media campaign (newspapers, radio stations etc.) to send a copy of their outgoing text messages (all or a selection) to a free number provided by *Swisscom* and connected to an automatic collection tool. All the text messages received were integrated into the corpus, which contains 25,947 text messages or about 500,000 words. 2,784 people participated in the study and half of them also filled in an online questionnaire. As these respondents were among the most productive, at least 75% of the messages can be linked to socio-demographic data.

The corpus contains the following language varieties (by decreasing order of number of text messages): Alemannic dialects (10,706), non-dialectal German (7,287), non-dialectal French (4,619), standard Italian (1,471), Romansh (1,121) and some other languages without official status in Switzerland. For the present study, we analyzed a sub-corpus consisting of 1,100 or about a quarter of the French messages. In this sub-corpus, we found 3,451 subjects, out of which 2,456 are clitics. The latter constitute the reference value for the following statistical analysis because dropped arguments have to be analyzed as pronominals (and, more precisely, as clitic pronouns in French; cf. Sigurðsson 2011:289).

4 Results

This section presents some statistical results with regard to subject omission in French text messages. We will see that subject omission in French text messages behaves mostly as in English diaries (cf. Haegeman, 2013 and this volume).

Table 1 shows the figures for clitic subject realization and clitic subject drop.

person		Σ	1sg	2sg	3sg referential	3sg expletive	1pl	2pl	3pl
clitic	N	2,456	1,191	496	533	153	15	49	19
	%	91.54	90.64	95.57	97.8	70.51	78.95	100	95
zero	N	227	123	23	12	64	4	0	1
	%	8.46	9.36	4.43	2.2	29.49	21.05	0	5
Sum (clitic + zero)	N	2,683	1,314	519	545	217	19	49	20
	%	100	100	100	100	100	100	100	100

Table 1: Subject drop per person in French text messages

As can be seen, the numbers for plural subjects are very low overall and therefore will not provide reliable data. They will not be taken into account in the following discussion.

There is however a significant difference between the omission of the 1st, 2nd, and 3rd referential person, respectively ($\chi^2 = 37.07$, $df = 2$, $p < 0.001$).² Table 1 indicates that the first person is omitted more frequently than the other two. The difference between 3rd referential person and 3rd expletive is also highly significant ($\chi^2 = 128.7421$, $df = 1$, $p < 0.001$). In other words, expletives are more likely to be dropped than referential subjects. This is illustrated in example (6) **Fehler! Verweisquelle konnte nicht gefunden werden.**, where the sentence initial *il* of *il faut* ‘it is necessary’ is dropped.

- (6) *Pas de souci ;-) ouais Ø faut qu'on se*
 (II)
 not of worry ;-) yes (It) is.necessary.PRS.3SG that=one.IMPERS.3.SG REFL.3.SG

fasse un truc bientôt!
 make.PRS.CONJ.3.SG a thing soon

 ‘No worries ;-) Yeah...(we) have to put up something together soon!’

Compared to other relevant empirical studies on subject drop in abbreviated registers,³ Table 2 shows that the overall omission rate in our corpus is quite parallel to, though slightly lower than in most of the diaries studied by Haegeman (1999; with Plath 1959 stepping out of line):

Source (cf. Haegeman 1999:131)	% of subject omission in root clauses
Plath (1959) English	25.92
Léautaud (1933) French	11.54
Woolf (1940) English	11.22
<i>sms4science</i> (2010) French	8.46

Table 2: Overall rate of subject drop in diaries vs. French text messages

4.1 “Avoid a”?

If subject omission in SMS were due to Weir’s (2012a) “Avoid weak start” principle, we should find not only argument omission, but also other kinds of prosodically caused omissions in our data which do not necessarily range over whole constituents. The absence of such data does, not, however, falsify the principle in question, as it could be due to chance or some other extra-linguistic factors influencing the data in our SMS corpus.

One possibility to test if omission is due to phonological principles is to look at article omission, as articles are by definition weak phonological elements (cf. Stowell 1991, 2013, Oosterhof & Rawoens and Reich, this volume.) Bare nouns in subject position are, however, very rare in our corpus. Overall, we find only 10 NPs opposed to 171 DPs and 75 QPs in subject position. 8 out of the 10 NPs do occur phrase-initially, but the tendency to be phrase-initial is also true for DPs and QPs in subject position. So there is not enough evidence to confirm (or refute) the “Avoid weak start” principle; a larger corpus would maybe provide a higher number of occurrences to make statistical tests possible or, alternatively, show more convincingly that article omission is in fact extremely rare in French text messages (cf. Reust 2015).

Another hint of a phonologically driven strategy of omission is word truncation at the beginning (e.g. of the message). Examples (7) to (9) are an exhaustive list of the message-initial word truncations we found in our sub-corpus.

- (7) *tain mon gars j'suis tombé sur un son electro de fou*
 (putain)
 whore my guy I=am fallen on a sound electro of crazy

‘Holy shit, buddy, I have come across a crazy electronic sound’

- (8) *mme parti pour la Tchequie*
 (sommme)
 have.1.PL left for the Czech Republic

‘We have left for the Czech Republic’

- (9) *scus Ça à duré plus long que prévu*
 (excuse)
 sorry that has lasted more long than expected

‘Sorry, that lasted longer than expected’

Furthermore, we did not find any example of auxiliary or copula omission with an inverted clitic subject present, i.e. parallel cases to (10) do not exist in our sub-corpus:

- (10) Ø You seen Tom? (English)
 (Have)

‘Have you seen Tom?’ (cf. Weir 2012a:109)

For complex verb forms, it appears that if the copula or auxiliary is dropped together with the subject (i.e. there is only the predicative complement or the participle left, cf. (11), see also example (4) in section 2) this will most likely occur at the beginning of an intonational unit (cf. Table 3, last column).

- (11) *Coucou! Ø Ø Bien reposée!?*
 (Es-tu)
 hello (are you) well rested-F.SG

‘Hi! (Are you) well rested!?’

Dropping of the subject while maintaining the copula or auxiliary shows the same tendency, but this preference for utterance-initial position is not as clear-cut as for subject plus copula/auxiliary drop.

subject type	null subject, aux/cop present		null subject, null cop/aux	
	N	%	N	%
beginning of intonational unit	112	75.17	72	92.31
within intonational unit	37	24.83	1	1.28
others	-	-	5 gapping	6.41
total	149	100	78	100

Table 3: Subject drop and auxiliary/copula drop

All in all, unambiguous evidence of a purely prosodic constraint (such as article drop in initial subjects or word truncations) is scarce, as the cases of subject (and auxiliary or copula) drop can also be explained as syntactic phenomena. The absence of evidence for the “Avoid weak start” principle is, of course, not in itself a proof or falsification of the underlying hypothesis; all we can state at the moment is that the available evidence is too weak to test it and that more ample text messaging data and similar data are needed.

4.2 ‘Reduced CP hypothesis’

According to Haegeman’s reduced CP hypothesis, subject omission should not be found in embedded sentences. Table 4 confirms this prediction by and large.

Subject types in embedded clauses	N	%
clitic	478	98.56
zero with <i>ya</i>	6	1.24
other zero	1	0.2
total	485	100

Table 4: Subject drop in embedded clauses

Out of 485 subjects in embedded contexts, only 7 are omitted. 6 of them are expletives in the presentative *il y a* (‘there is’). In spoken French, the fusion of the colloquial allomorph of *il* ([i]) and the subsequent phonetically identical *y* ([i]) is common (cf. Blanche-Benveniste 2000:38). It is therefore probable that the spelling in our data imitates this phonetic realization (“phonetic spelling”): [ilja] > [ija] > [ja] at the phonetic level corresponds to the graphic realization <ya> for the normative <il y a>.

The remaining example is most probably a typing error (cf. (12)); <le>, ‘it’, instead of intended <je>, ‘I’; <l> and <j> are on the same key on older mobile phones, and the position of the direct object clitic *le* in (12) is odd or even ungrammatical (it should be in front of its governing verb, *écrire*, ‘write’).

- (12) *Mais tu peux juste me dire keske tu à fais ce*
 but you can.1/2.SG just me.DAT tell.INF what you have.2.SG done this
week pour ke Ø le puisse déjà écrire
 week, so that I it.ACC can.PRS.CONJ.1./3.SG already write.INF,
 ‘But you can just tell me what you did this week so that I can already write it down.’

In inverted contexts and with fronted arguments, subject omission should not be possible either according to the ‘reduced CP hypothesis’. We find 61 clitics in inversion and no clear case of omission in these contexts.⁴

For questions, there are some cases of subject drop with *ya*, which we consider, however, as the phonetic spelling of the agglutinative *il y a* ([ija]) in colloquial French (see above), with *wh in situ* (cf. (13)), and 6 examples of *comment va* (‘how goes’) with a preceding modal complement/question word and still without the expletive demonstrative *ça* (cf. (14) and the next section for discussion). No other occurrences of subject drop after a fronted *wh*-element were found:

- (13) *si jpx venir avan pr vs aidez tu dis!*
 (je peux) (pour vous)
 if I=can.1./2.SG come.INF before for you.ACC.2.PL help.INF, you say.2.SG
 et Ø faut amener quoi?
 (il)
 and it is.necessary.3.SG bring.INF what

‘If I may come earlier to help you, please let me know ! And what would you like us to bring along ?’

(14) *Hey, comment Ø va? As-tu une adresse e-mail pour que je*
(ça)
 hey how it goes have.2.SG=you an address e-mail for that I
puisse te répondre?
 can.PRS.CONJ.1./3.SG you.DAT answer.INF

‘Hey, how’s it going? Do you have an email address so that I can answer you ?’

Table 5 shows the rates of clitic subjects and subject drop according to the preceding syntactic context.⁵

Subject type	clitic		zero (and <i>ya</i>)	
	N	%	N	%
no preceding element	1420	90.97	141	9.03
complements	83	93.26	6	6.74
adjuncts	342	97.16	10	2.84
coordinating conjunction	378	91.75	34	8.25
complementizer	484	99.79	1	0.21
interrogative particle (<i>est-ce que</i>)	24	100	0	0
doubled subject	57	95	3	5
total	2788	93.46	195	6.54

Table 5: Subject drop and preceding elements

A few additional comments are due here. The omission after a complementizer is illustrated in example (12) above and most probably a typing (performance) error. We do find fronted adjuncts with dropped subjects (cf. example (15)) as well as coordinating conjunctions (cf. (16)) with subject omission, which is expected according to Haegeman:

(15) *I am overbooked jusko vendredi soir: et de jour Ø suis*
(je)
 I am overbooked till=at.the Friday evening: and of day (I) am
en stage à la radio !
 at training.course at the radio!

‘I am overbooked until Friday evening : and during the day, I attend a training course at the radio!’

(16) *Ouimais Ø suis encore très fatiguée et de plus en plus de*

(je)
 ‘Yes.but (I) am still very tired-F.SG and of more and more of
cernes... vivement les résultats! becs
 dark.circles.under.the.eyes... can’t.wait.for the results! Kisses

‘Yes, but (I) am very tired, more and more dark circles under my eyes...Can’t wait
 for the results ! Kisses’

What, on the other hand, is not expected under Haegeman’s ‘reduced CP analysis’ with SubjP as the root phase are the three cases we found of a strong (topicalized?) subject pronouns not followed by a clitic subject, as exemplified in (17).

(17) *Bizarre. En plus elle m' avait dit qu'elle pouvait pas*
 strange In addition, she me.DAT had.3.SG said that she could.3.SG NEG

en fin d' aprem car elle retournait sur lausanne.
 in end of=afternoon for she returned.3.SG to Lausanne

Bref ben moi mtn Ø me suis organisé différemment
 (bien) (maintenant) (je)
 short well me.TOP now, (I) REFL.1.SG am organized differently

‘Strange. Also, she had told me that she was not available in the afternoon, because she would return to Lausanne. In short, I have organized myself differently now.’

This is clearly a case of a fronted argument (subject) with subject drop, which should not be possible (no landing side for the moved topic available) under the ‘reduced CP hypothesis’. We will come back to these examples in the discussion section.

Finally, it should be mentioned that our corpus contains in general very few cases of clitic left dislocations (topic-shifting structures as in (18)) and cleft sentences (focus-marking structures, cf. (19)). Only 55 out of 520 subjects are doubled by a clitic (which amounts to 10.58%, vs. Culbertson’s (2010:116) analysis of the spoken *Lyon* corpus with 80.6% and the *PFC* corpus with 56.3%), and in our data, there are 22 cleft sentences found (in the whole corpus only 55, cf. Stark 2014).

(18) *Ouais la soirée MacFlurry ça me plait!*
yeah the evening MacFlurry that me.DAT please.3.SG

‘Yeah, the MacFlurry evening, I really like that!’

(19) *Coucou! c'est pas aujourd'hui qu'on va voir ce film à l'abc?*
hi! It=is NEG today that=one goes see.INF that movie at the=abc

c'est à quelle heure?
it=is at what hour

‘Hi ! Isn’t it this evening that we will go to watch this movie at the abc ? Which time ?’

So there is some additional evidence for an inactive or absent left periphery and a reduced communicative need for informationally marked structures in text messages. We will try to interpret these results in the next section.

5 *Discussion*

As Table 1 and Table 2 illustrate, about 10% of subjects are dropped in our corpus, mostly 1st and 3rd person singular clitics (with a strong preference for dropping expletives). There is not, however, much evidence in favor of a generalized “Avoid weak start” constraint (see section 4.1. above), as we did not find significant results for initially dropped articles or internal word truncation or only auxiliary-drop. There might be some preference for the typical French “Accentual Phrase” (cf. Jun & Fougeron 2000) with none or only one weak syllable before the first high tone/accent (see example (4)) and the tendency to drop the clitic subject *and* the weak auxiliary/copula sentence-initially, but this cannot be stated with certainty.

Much stronger evidence was found for the ‘reduced CP hypothesis’. Subject drop in embedded sentences is virtually absent if we discard the *ya-* examples as cases of phonetic spelling (of *il y a* ‘there is’) and example (12) as a typing error. The adjunct/argument asymmetry holds completely, if we discard *comment va?* (for *comment ça va* ‘How’s it going’) and *moi* (‘me’) without the coreferential clitic *je* (‘I’) for the moment (cf. Haegeman 1997, 1999, 2013). Additionally, there are few examples with an unfolded active information structure (cf. rare dislocations, clefts) in the corpus in general. Coming back to the *comment va-*structures, we do not have any evidence for subject drop with moved *wh*-elements, except these 6 examples, which represent thus the first problematic data for a register-specific ‘reduced CP’ account of subject drop. We are in all likelihood dealing here with fully lexicalized structures without any compositional syntax, however, similar to Austrian German *Wie geht?* (for *Wie geht’s* ‘How’s it going?’ without the expletive pronoun *es* ‘it’) or, *How is going?* without the expletive pronoun *it* in some colloquial varieties of English.

Much more intriguing for the ‘reduced CP hypothesis’ are fronted subject arguments (*moi*) with or without following adverbial adjuncts and no coreferential clitic subject in its canonical position (see example (17) above). *Moi* ‘me’ is the prosodically strong allomorph of the nominative clitic *je* (‘I’) and the accusative/dative clitic *me* (‘me, ‘to me’), which occurs obligatorily in stressed positions such as isolated answers to *wh*-questions (*Qui est arrivé? – Moi/*Je*; ‘Who has arrived? - Me?’), after prepositions (*avec moi/*avec je/me*, ‘with me’) or in focus position (*c’est moi qui pense cela/*c’est je qui pense cela* – ‘It is me who thinks that’). The standard French equivalent of example (17) contains a doubling structure *moi – maintenant – je* (‘me, now I’), which is said to be a topic-marking (topic-shifting) device in French, *Clitic Left Dislocation* (CLLD, cf. e.g. Cinque 1990, Haegeman 2004). The analysis of all three examples of *moi* without *je* in our data clearly indicates an informationally marked structure in each case, with *moi* serving as a new, or at least contrastive, topic. In example (17) above, *moi* indicates a clear topic change. Examples (20) and (21) below show a

contrastive topic in the sense of Mayol (2010). (20) can be analyzed as an implicit contrast between the newly established topic *moi* and the preceding one (between the pronoun's referent "and the other entity salient in the discourse", Mayol 2010:2501). Note that the same new topic (*moi*) is subsequently mentioned with a resumptive *je* in the same message. (21) finally is qualified as a case of weak contrast between *moi* and some other discourse referent, where "the speaker is only making a claim about the referent of the OSP [overt subject pronoun, ES & ART] and leaves open whether this claim should or should not apply to the other referents relevant in the discourse" (Mayol 2010: 2501f.).

(20) *HeY! Thanks a lot pr ton message! Moi aussi,*
 hey! thanks a lot for your message! Me.TOP too,

\emptyset *m réjoui 2 tomoro!*
 (Je)
 (I) REFL.1.SG look.forward.1.SG of tomorrow

'Hey! Thanks a lot for your message! Me too, I look forward to tomorrow!'

(21) *Ça c I bonne nouvelle au réveil! \emptyset Suis bien contente ke*
 that this=is a good news at=the waking.up (Je) \emptyset am very happy that

Ça s' arrange finalement ! Moi \emptyset ai reÇu un mail
 that REFL.3.SG settles.up finally me.TOP \emptyset have.1.SG received a message
de Ashleigh mercati qui veut qu'on se revoit
 from Ashleigh mercati who wants that=one REFL.3.SG see.again.3.SG

'That is really good news at the moment of waking up! (I) am really happy that things are finally working out! As for me, I have received an email from Ashleigh Mercati who wants us to meet again'

These findings with clearly topical *moi* (in TopP or some other functional projection in the left periphery) make it impossible to assume a truncated CP (from FinP up to ForceP, including FocP and TopP) above SubjP for our data, as assumed in Haegeman (2013 and this volume) for diary subject omission and similar types of Written Subject Omission (WSO). An

analysis of the whole French corpus reveals 10 cases from 9 different native speakers of *moi* without *je* (cf. Robert-Tissot forthcoming). Even if this is not a frequent pattern, it is nonetheless recurring and thus seems to be productive. One way out of this problem could be the assumption that *moi* without *je* is a kind of ‘written free allomorph’ to simple *je*, situated in the canonical subject position, SpecSubjP or SpecTP. Accordingly, examples (17), (20) and (21) would not be cases of subject omission at all. Based on some examples of *moi je*-recordings with a prosodically extremely reduced *je* [mwaʒ], Detges (2013) assumes in a similar vein that *moi-je* with a strongly reduced second element forming the syllable coda with a final voiced fricative is on its way to becoming a free allomorph of *je* in colloquial French. Contrary to his examples, however, ours are cases of informationally motivated *moi* (topical), and such elements are traditionally analyzed as being located higher in the left periphery than subjects (cf., e.g., Frascarelli & Hinterhölzl 2007). If this analysis is on the right track, we may conclude that topics can be moved to the left periphery without obligatorily being resumed by a clitic in the proposition. Instances of subject drop could then be analyzed as silent familiar topics (cf. Frascarelli & Hinterhölzl 2007; see Robert-Tissot forthcoming for further details).

Except for the *moi*-structures without coreferential *je*, the ‘reduced CP hypothesis’ for Written Subject Omission (WSO) of Haegeman (2013 and this volume) can be fully applied to subject omission in our corpus of French text messages.

6 Conclusion

We have discussed in this contribution one specific syntactic property of text messages, i.e. subject omission, also in cases where it would be ungrammatical in spoken varieties of French (e.g. with preposed adjuncts). Text messages represent a graphic abbreviated register with two specific properties: they show a close-to-phonetics spelling and a considerable degree of interactivity, compared, e.g., to e-mails. Searching for an

adequate analysis of our findings, we tested the predictions made by the postsyntactic prosodic account of Weir (2012a) for English subject omission, but found that, while there might be some indication of the prosodic preferences of spoken French reflected in our data (subject clitic and auxiliary/copula drop CP-initially), the evidence for the ‘reduced CP hypothesis’ put forward in Liliane Haegeman’s work and also Weir (2012b) is more robust. This hypothesis is compatible with the specific, reduced interactionality of text messages compared to real face-to-face interaction, which is due to their asynchronous, graphic character. Accordingly, the frequencies of informationally marked syntactic structures such as (clitic) left dislocations or cleft sentences are very low, maybe because in the default case, deictic anchoring of the utterance to the immediate context is sufficient (as in recipes, headlines, diaries, notes etc.), and an articulated CP is simply not necessary in any instance of a text message. Some instances of clitic subject drop with stressed topical *moi* (‘me’) represent, however, a serious challenge to this account, as subjects seem to be dropped here after preposed arguments (in SpecTopP). This points to another possible analysis of (referential) subject drop in abbreviated written registers, rejected by Haegeman, but maybe promising at least for referential subjects, i.e. familiar topic drop (cf. Robert-Tissot forthcoming).

References

Corpora:

Lyon Corpus, CHILDES: <http://childes.psy.cmu.edu/data/Romance>

PFC (Phonologie du Français Contemporain): <http://www.projet-pfc.net>; see:

Durand, Jacques, Bernard Laks & Chantal Lyche. 2002. La phonologie du français contemporain: usages, variétés et structure. In C. Pusch & W. Raible (eds.):

Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance

Corpus Linguistics - Corpora and Spoken Language, 93-106. Tübingen: Gunter Narr Verlag.

Durand, Jacques, Bernard Laks & Chantal Lyche. 2009. Le projet PFC: une source de données primaires structurées. In J. Durand, B. Laks & C. Lyche (eds.): *Phonologie, variation et accents du français*, 19-61. Paris: Hermès.

sms4science.ch Corpus: <http://www.sms4science.ch>; see:

Stark, Elisabeth, Ruef, Beni & Simone Ueberwasser. 2009-2014. *SMS Corpus*. University of Zurich. <https://sms.linguistik.uzh.ch>

Scientific literature:

Androutsopoulos, Jannis & Gurly Schmidt. 2002. SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe. *Zeitschrift für Angewandte Linguistik* 36. 49-79.

Anis, Jacques. 2007. Neography – Unconventional Spelling in French SMS Text Messages. In B. Danet & S. C. Herring (eds.), *The Multilingual Internet – Language, Culture and Communication Online*, 87-115. New York: Oxford University Press.

Béguelin, Marie-José. 2012. La variation graphique dans le corpus suisse de SMS en français. In S. Caddéo, et al. (eds.), *Penser les langues avec Claire Blanche-Benveniste*, 47-63. Aix-en-Provence: Presses de l'Université de Provence.

Biber, Douglas. 1995. *Dimensions of Register Variation*. Cambridge: Cambridge University Press.

Bieswanger, Markus. 2013. Micro-linguistic structural features of computer-mediated communication. In S. Herring, D. Stein & T. Virtanen (eds.), *Handbook of the Pragmatics of Computer-Mediated Communication*, 463-485. Berlin & New York: Mouton de Gruyter.

Blanche-Benveniste, Claire, 2nd edn. 2000. *Approches de la langue parlée en français*. Gap: Ophrys.

Cardinaletti, Anna. 2004. Toward a cartography of subject positions. In L. Rizzi (ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures Volume 2*, 115–165. Oxford, Oxford University Press.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1-52. Cambridge & Mass.: MIT Press.

Cinque, Guglielmo. 1990. *Types of A'-Dependencies*. Cambridge, Mass.: MIT Press.

Culbertson, Jennifer. 2010. Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86(1). 85-132.

- Dell, François. 1984. L'accentuation dans les phrases en français. In F. Dell, D. Hirst & J.-R. Vergnaud (eds.), *Forme sonore du langage: Structure des représentations en phonologie*, 65-122. Paris: Hermann.
- Delais-Roussarie, Elisabeth. 1996. Phonological Phrasing and Accentuation in French. In M. Nespore & N. Smith (eds.), *Dam Phonology*, 1-38. La Haye: Holland Academic Graphics.
- Detges, Ulrich. 2013. First person pronouns in spoken French. A case study in cliticization. In K. Jeppesen Kragh & J. Lindschouw (eds.), *Deixis and Pronouns in Romance Languages*, 33-47. Amsterdam & Philadelphia: Benjamins.
- Fairon, Cédric, Jean René Klein & Sébastien Paumier. 2006. *Le Langage SMS. Étude d'un corpus informatisé à partir de l'enquête « Faites don de vos SMS à la science »*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Ferguson, Charles A. 1982. Simplified registers and linguistic theory. In L. K. Obler & L. Menn (eds.), *Exceptional Language and Linguistics*, 49-66. New York: Academic Press.
- Frascarelli, Mara & Hinterhölzl, Roland. 2007. Types of topics in German and Italian. In K. Schwabe & S. Winkler (eds.), *On Information Structure, Meaning and Form. Generalizations across languages*, 87-116. Amsterdam & Philadelphia: Benjamins.
- Haegeman, Liliane. 1997. Register variation, truncation and subject omission in English and in French. *English Language and Linguistics* 1. 233-270.
- Haegeman, Liliane. 1999. Adult null subjects in non *pro*-drop languages. In M.-A. Friedemann & L. Rizzi (eds.), *The Acquisition of Syntax*, 128-169. London: Addison, Wesley and Longman.
- Haegeman, Liliane. 2004. Topicalization, CLLD and the Left Periphery. *ZAS Papers in Linguistics* 35. 157-192.
- Haegeman, Liliane. 2013. The syntax of registers. Diary subject omission and the privilege of the root. *Lingua* 130. 88-110.
- Haegeman, Liliane. (this volume). Unspeakable sentences. Subject omission in written registers: a cartographic analysis.
- Haegeman, Liliane & Tabea Ihsane. 1999. Subject ellipsis in embedded clauses in English. *English Language and Linguistics* 3, 117-145.
- Haegeman, Liliane & Tabea Ihsane. 2001. Adult Null Subjects in the non-*pro*-drop Languages: Two Diary Dialects. *Language Acquisition* 9. 329-346.

- Herring, Susan C. 2007. A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet* 4, article 1, <http://www.languageatinternet.org/articles/2007/761>. (24 July, 2015.)
- Hymes, Dell. 1971. *Pidginization and creolization of language*. Cambridge: Cambridge University Press.
- Janda, Richard J. 1985. Note-taking English as a simplified register. *Discourse Processes* 8. 437-454.
- Jun, Sun-Ah & Fougeron, Cécile. 2000. A Phonological Model of French Intonation. In A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology*, 209-242. Dordrecht: Kluwer Academic Publishers.
- Little, Greta D. 1978. Subject deletion in English. *Linguistic Association of Canada and the United States: LACUS forum* 5. 289-297.
- Mayol, Laia. 2010. Contrastive pronouns in null-subject Romance languages. *Lingua* 120. 2497–2514.
- Napoli, Donna Jo. 1982. Initial material deletion in English. *Glossa* 16(1). 85–111.
- Nariyama, Shigeko. 2004. Subject ellipsis in English. *Journal of Pragmatics* 36. 237–264.
- Nariyama, Shigeko. 2013. Discourse Coherence and Referent Identification of Subject Ellipsis in Japanese. In A. Włodarczyk & H. Włodarczyk (eds.): *Meta-Informative Centering in Utterances: Between Semantics and Pragmatics*, 167-181. Amsterdam: Benjamins.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Oh, Sun-Young. 2005. English zero anaphora as an interactional resource, 1, *Research on Language and Social Interaction* 38. 267-302.
- Oh, Sun-Young. 2006. English zero anaphora as an interactional resource, 2, *Discourse studies* 8. 817-846.
- Oosterhof, Albert & Gudrun Rawoens. (this volume). Register variation and distributional patterns in article omission in Dutch headlines.
- Panckhurst, Rachel. 2009. Short Message Service (SMS). Typologie et problématiques futures. In T. Arnavielle (ed.), *Polyphonies, pour Michelle Lanvin*, 33–52. Montpellier: Université Paul Valéry Montpellier 3.
- Reich, Ingo. (this volume). On the omission of articles and copulae in German.
- Reust, Fabiola. 2015. *L'ellipse de l'article dans les SMS suisses français*. Zurich: University of Zurich MA thesis.

- Rizzi, Luigi. 1997. The fine structure of the left periphery. In L. Haegeman (ed.), *Elements of Grammar*, 281-337. Dordrecht, Boston & London: Kluwer.
- Rizzi, Luigi & Shlonsky, Ur. 2007. Strategies of subject extraction. In H-M. Gärtner & U. Sauerland (eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, 115–60. Berlin: Mouton de Gruyter.
- Robert-Tissot, Aurélia. Forthoming. *Le sujet et son absence dans les SMS français. Une analyse syntaxique*. Vincennes: Presses universitaires de Vincennes.
- Sigurðsson, Halldór Á. 2011. Conditions on Argument Drop. *Linguistic Inquiry* 42(2). 267-304.
- Stark, Elisabeth. 2014. Frequency, Form and Function of Cleft Constructions in the Swiss SMS Corpus. In A.-M. de Cesare (ed.), *Frequency, Form and Function of cleft constructions*, 325-344. Berlin, Munich & Boston: Mouton de Gruyter.
- Stark, Elisabeth. 2015. 'De l'oral dans l'écrit'? – Le profil variationnel des SMS (textos) et leur valeur pour la recherche linguistique. In K. Jeppesen Kragh & J. Lindschow (eds.), *Les variations diasystématiques et leurs interdépendances dans les langues romanes. Actes du Colloque DIA II à Copenhague (19-21 nov. 2012)*, 395-405. Strasbourg: Editions de linguistique et de philologie.
- Stowell, Tim (1991) "Empty Heads in Abbreviated English," GLOW Abstract, GLOW Newsletter #26, HAG.
- Stowell, Tim (2013) "The Syntax of Abbreviated English," lecture given at the International Congress of Linguists, Geneva, July 22, 2013.
- Teddiman, Laura. 2011. Subject ellipsis by text type: an investigation using ICE-GB. In J. Newman, H. Baayen & S. Rice (eds.): *Corpus-based studies in language use, language learning, and language documentation*, 71-88. Amsterdam: Rodopi.
- Teddiman, Laura & John Newman. 2007. Subject Ellipsis in English: Construction of and Findings from a Diary Corpus, *26th conference on Lexis and Grammar, Bonifacio, 2-6 October 2007*. Alberta: University of Alberta: Ms.
- Thurlow, Crispin & Michèle Poff. 2013. Text Messaging. In S. C. Herring, D. Stein & T. Virtanen (eds.), *Handbook of the Pragmatics of Computer-Mediated Communication*, 163-190. Berlin & New York: Mouton de Gruyter.
- Walther, Joseph B. 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research* 23. 3-43.

- Weir, Andrew. 2009. Subject pronoun drop in informal English. Richard M. Hogg Prize winning essay. <http://people.umass.edu/aweir>. (24 July, 2015.)
- Weir, Andrew. 2012a. Left edge deletion in English and subject omission in diaries. *English Language and Linguistics* 16(1). 105-129.
- Weir, Andrew. 2012b. Article drop in headlines. Talk given at ECO5, 7th of April 2012. Ms.
- Weir, Andrew. (this volume). Object drop and article drop in reduced written register.
- Zdorenko, Tatiana. 2010. Subject omission in Russian: a study of the Russian National Corpus. In St. Th. Gries, St. Wulff & M. Davies (eds.): *Corpus-Linguistic Applications: Current Studies, New Directions*, 119-133. Amsterdam: Rodopi.

Endnotes:

¹ See Little (1978) for a descriptive corpus-based study on subject ellipsis in informal written English.

² The chi-square test was applied in order to test whether the difference between 1st, 2nd and 3rd person is due to chance or significant. The higher the χ^2 score the more significant the result (depending on the degrees of freedom). The degrees of freedom (df) indicate how values or independent pieces of data are used to make a calculation. If $p < 0.001$, the probability that the distribution is accidental is lower than 0.1% (the lower p, the more significant).

³ Note that to our knowledge, no empirical work on subject drop in sufficiently large corpora of text messages or CMC (for languages without systematic argument drop) in general is available to date, so we have to compare our results to monological abbreviated registers such as diaries.

⁴ It must be mentioned that clear cases of subject omission in inversion are difficult to identify because French has the possibility of marking questions by intonation without any word order modification. So, it will not be possible to decide if an omitted subject in questions should be analyzed as being inverted or not, except with the special form *puis-je*, where the indicative form is replaced by a special morphological form.

⁵ There can be more than one constituent preceding the subject. In this case, the subjects are counted for each context.