

Article

The Role of Functional Heads in Code-Switching Evidence from Swiss Text Messages (sms4science.ch)

Aurélia Robert-Tissot ^{1,*} and Etienne Morel ^{1,2}

¹ Romanisches Seminar, University of Zurich, Zürichbergstrasse 8, CH-8032 Zürich, Switzerland

² Institut des sciences du langage et de la communication, University of Neuchâtel, Avenue DuPeyrou 6, CH-2000 Neuchâtel, Switzerland; etiennejoseph.morel@uzh.ch

* Correspondence: aurelianatascha.robert-tissot@uzh.ch

Academic Editors: Ji Young Shim and Maria del Carmen Parafita Couto

Received: 28 November 2016; Accepted: 5 July 2017; Published: 13 July 2017

Abstract: This study aims to test two principles of code-switching (CS) formulated by González Vilbazo (2005): The Principle of the Functional Restriction (PFR) and the Principle of Agreement (PA). The first states that a code-switch between the morphological exponents of functional heads belonging to the same extended projection of a lexical category (N° or V°) is not possible. The second claims that inside a phrase, agreement requirements have to be satisfied, regardless of the language providing the lexical material. The corpus on which we tested these hypotheses consists of 25,947 authentic text messages collected in Switzerland in 2009 and 2010. In our corpus, the PA is maintained. The PFR also seems to hold, even if data is limited. Interestingly, contradicting examples can be explained by phonological principles or the sociolinguistic background of the authors, who are not native speakers. Overall, the evidence found in spontaneously written non-standard data like text messages seems to confirm the validity of the two principles.

Keywords: code-switching; SMS; syntactic principle; functional restriction; agreement

1. Introduction

In this contribution, we present the results of a pilot study on the syntactic restrictions of code-switching (CS). More precisely, we will test two principles proposed by González Vilbazo [1] on a corpus of genuine text messages collected in Switzerland. González Vilbazo's principles build on previous generative analyses of CS, to which he integrates the results of his analysis carried out on a corpus of spoken interaction between young Spanish-German bilinguals [1]. We will see that his principles are not only valid for spoken, but also for written informal interaction, at least for switches between and within Germanic (English, German) and Romance (French, Spanish, Italian) languages. This is a relevant observation, as it indicates that the principles may in fact guide the syntactic structure of CS for a larger variety of languages and independently from the extralinguistic context in which CS occurs.

In what follows, we first present some of the challenges faced when approaching CS and make explicit our approach to CS in this paper (Section 1.1). Section 1.2 gives a short outline of existing approaches to CS and grammar, followed by a detailed presentation of the principles tested in Section 1.3, focusing on the differences between González Vilbazo's approach [1] and the alternative proposals. In Section 1.4 we briefly review existing studies on CS in texting. The data and methods used in this study are presented in Section 2. Section 3 deals with the results of the analysis and discusses them. The conclusion in Section 4 addresses some limitations of this study and provides an outlook for further research.

1.1. Approaching Code-Switching

Since its first use by Vogt in 1954 [2], the term CS was adopted by different fields in linguistics. Following Gumperz' much cited definition, CS can be described as the "juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems" [3] (p. 59). However, depending on the theoretical approach (formal, sociolinguistic, interactional etc.), the unity of analysis in question (sentence, turns in talk, sequences of interaction), and sometimes on the specific extralinguistic context, what can all be referred to as CS, can prove to be very differing phenomena (see [4,5] for a discussion). In fact, both facets of the term CS (*code* and *switching*) raise important questions as to their precise definition.

Firstly, it is not straightforward to define without ambiguities what a "code" exactly is. It is a subject to debate whether a given linguistic material can always be attributed to a labeled language or whether it can be legitimately assumed to belong to different languages or varieties of languages [6–8]; the type of data studied here (written, mainly informal interaction) and the social activities that are often conducted in this context (playful language use, see Section 1.4) may even amplify these challenges [7,8].

Secondly, it is not evident what represents an actual instance of switching, even if the codes are unambiguously labelled: Which criteria (sociopragmatic, morphological, etymological/lexicographic) and which perspective (the researcher's or the participants') is more appropriate to identify a switch?

Thirdly, the mere impossibility of finding a widely accepted definition of CS goes hand in hand with the difficulty in distinguishing CS from other linguistic contact phenomena, especially from borrowing [4,9]. The lack of a clear-cut distinction leaves us with a continuum ranging from clear instances of CS to clear instances of borrowings shared by a larger linguistic community. In the corpus of text messages under analysis (see Section 2), many allogenic items appear to be neither used completely spontaneously, nor to be really integrated into the repertoire shared by a larger community, hence belonging to the grey zone of the continuum between the two categories [4,8,10].

These caveats being formulated, we underline the fact that the present paper focuses on the formal analysis of CS based on two syntactic principles formulated by González Vilbazo [1], which we test without any theoretical presupposition on what grammatical CS entails. It is based on the empirical identification of two or more constituents that expose the morphological material of what appears-to researchers-as different languages or varieties inside the relevant syntactic boundaries. This may also imply simple word insertions, which are not considered to be CS by any of the theoretical approaches (see [4] for a discussion).

1.2. Code-Switching and Grammar

Since the seminal work of Poplack [11], research on syntactic restrictions in CS has grown to a large body of works, which we will not be able to address in detail here; for the sake of brevity, we focus on the research that is particularly relevant to the principles tested, see Section 1.3.

In the generative framework, CS has become popular with Poplack's proposal of an Equivalence Constraint on CS [11]. However, this constraint stating that parallel structures in two languages are needed in order to switch from one language to the other¹ has repeatedly been shown not to hold up to scrutiny (cf. [1] (p. 30) and [8] (pp. 40–54)). Other constraints have been proposed to overcome the shortcomings of the Equivalence Constraint, such as Belazi et al.'s Functional Head Constraint (FHC) [12]. Instead of focusing on the surface structure, the FHC takes into account the syntactic structure of phrases; it supposes that it is impossible to switch language between a functional head and its complement because of the strong relation between the two constituents. Yet, this would

¹ Code-switches tend to occur at points in discourse where the juxtaposition of L1 and L2 elements does not violate a syntactic rule of either language, i.e., at points around which the surface structure of the two languages map onto each other [11] (p. 586).

imply that CS is not possible between a determiner and its complement, as González Vilbazo points out [1] (p. 41), a fact that is not confirmed in his corpus, neither by our examples, as we will see in Section 3.2. Then, MacSwan's minimalist approach is based on the selectional properties of lexical items [13], since according to early minimalism, the relevant linguistic properties are all encoded in the lexicon. As a consequence, CS would only be constrained by feature mismatches of lexical items during computation. On this account, CS should be possible regardless of the heads involved, an assumption that does not seem to hold either. For instance, a relative pronoun must be followed by functional material of the same language, as shown by the following example taken from [1] (p. 79); the Spanish relative *que* 'who' has to be followed by the Spanish inflected verb *da* 'gives' (1a), while it is not possible to combine the German relative *der* 'who' with Spanish verbal morphology, be it in the Spanish (2b) or the German (2c) word order.

- | | | | | | | | |
|----|----|---|----------------------|-----------|-----------|-----------------------|-------------|
| 1. | a. | <i>El</i> Lehrer, | <i>que</i> <i>da</i> | schlechte | mündliche | Noten... ² | |
| | | the teacher | who gives | bad | oral | marks | |
| | b. | * <i>El</i> Lehrer, | <i>der</i> <i>da</i> | schlechte | mündliche | Noten | |
| | | the teacher | who gives | bad | oral | marks | |
| | c. | * <i>El</i> Lehrer, | <i>der</i> | schlechte | mündliche | Noten | <i>da</i> |
| | | the teacher | who | bad | oral | marks | gives |
| | | ‘The teacher who gives bad oral marks . . . ’ | | | | | [1] (p. 79) |

In this approach, it is not clear how different word order properties (cf. example 10 below) are dealt with.

The last model we want to mention is the Matrix Language Frame (MLF) model [14], of which the 4-M model, named after four distinctive morpheme classes, i.e. content morphemes and three types of system morphemes, whose characteristics play a crucial role when it comes to CS, (e.g., [15,16]) is a refined version. This model takes its origin in the assumption that there is a matrix language defined for a whole CP. Inside a CP, only the so-called “outsider late system morphemes” have to be in the matrix language (ML). This includes all functional morphemes showing a grammatical relation with constituents that are outside the immediate maximal projection, e.g., subject-agreeing morphemes on verbs. The MLF also allows for embedded language (EL) islands in which a whole chunk appears in the EL without being affected by the ML. This last component of the model is at the same time its weak point, since EL occurrences that do not confirm the expectations can be regarded as such islands. A more promising approach would be to propose alternative explanations for examples which, at first sight, do not seem to fit into the model.

Given the fact that the models discussed in this section seem to all have their downsides, we will test two principles more recently proposed by González Vilbazo [1], for which little research has been done so far. These principles are presented in more detail in the next section.

1.3. Presentation of the Principle of Agreement and the Principle of Functional Restriction

The principles we are going to test on our corpus are the Principle of Agreement (PA), and the Principle of Functional Restriction (PFR):

The PA states that:

All the morphosyntactic requirements of all the lexical and functional entities have to be satisfied in the sentence. Among these there is selection, agreement in the narrow sense,

² When there are different source languages in the examples, these are made evident by different fonts (e.g., recte vs. italics vs. bold).

and the attribution of case and theta roles. The language of the lexical entities does not matter, as long as they satisfy the requirements.³ [1] (p. 157, our translation)

The second principle, the PFR, is about where CS can occur, stating that:

Two functional heads X° and Y° have to be filled by lexical material of the same language if the functional category of YP is the complement of X° and both heads are part of the same extended projection.⁴ [1] (p. 94, our translation)

Because the PFR applies to extended projections, we focus our analysis on two extended projections, namely the extended projection of the noun (the DP) and the extended projection of the verb (CP). Extended projections form the functional overhead of the lexical categories. Hence, besides the lexical noun head N° , the DP contains at least a quantifier head Q° as well as a determiner head D° .⁵ Following González Vilbazo [1] (p. 108) and Bhatt [17], the DP can be represented as in Figure 1:

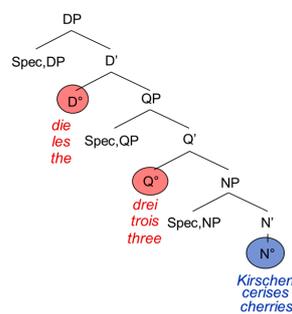


Figure 1. Simplified structure of the DP.

The determiner occupies the D° -head, the numeral the Q° -head, and the nominal expression the N° -head. It is not uncontroversial whether quantifiers have to be analyzed as heads or as phrases (or even as “semi-functional or semi-lexical categories” [18] (p. 408; for a general discussion, see Part III, Chap. 2)); for reasons that will become clear later, we follow [1], which considers them as heads. The DP can be modified by adjectives, which according to Cinque are phrases (occupying the specifier position of the adjective phrase) [19]. The adjective phrase is left-adjoined to the noun phrase as in Figure 2:

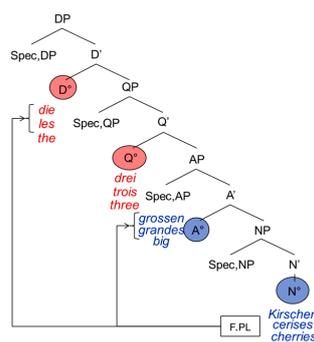


Figure 2. Simplified structure of the DP.

³ German original: *Alle morphosyntaktischen Anforderungen aller lexikalischen und funktionalen Einheiten müssen im Satz erfüllt sein. Dazu gehört Selektion, Kongruenz im engeren Sinne und das Zuweisen von Kasus und Thetarollen. Aus welcher Sprache die lexikalischen Einheiten stammen ist hierbei irrelevant, solange sie die Anforderungen erfüllen.*

⁴ German original: *Zwei funktionale Köpfe X° und Y° müssen lexikalisch mit Material aus derselben Sprache gefüllt sein, wenn die funktionale Kategorie YP Komplement von X° ist und beide Köpfe Teil des selben [sic] funktionalen Überbaus sind.*

⁵ Here we are simplifying the structure for the sake of clarity. Romance DPs for example also contain at least a Gender and a Case head cf. [19] (Part II, Chap. 3).

Coming back to the PA, the prediction is that inside, the DP agreement must be preserved regardless of the lexical item’s origin language. So, in Figure 2 we can see the feminine plural agreement on both the German and the French determiner as well as on the adjectives, triggered by the feminine plural *cherries* (*Kirschen* and *cerises*, respectively). It would be ungrammatical not to have agreement on lexical items that usually display it:

- | | | | | |
|----|--------------|-------|---------------|----------|
| 2. | die | drei | gross*(e)*(n) | Kirschen |
| | the.PL | three | big(F)(PL) | cherries |
| 3. | les/*la | trois | grand*(e)*(s) | cerises |
| | the.PL/the.F | three | big(F)(PL) | cherries |

Taking the determiner and the quantifier to be functional heads, a CS from the determiner to the quantifier is banned according to the PFR, since they are two functional heads of the same extended projection. The adjectival lexical item on the other hand, can originate from a different language, since it constitutes a lexical phrase:

- | | | | | |
|----|--------|------------|-----------------|------------------|
| 4. | die | drei/trois | grossen/grandes | Kirschen/cerises |
| | the.PL | three | big.F.PL | cherries |

The second extended projection we look at is the complementizer phrase CP, the functional overhead of the verb. The CP consists of the lexical verb V° , the “functional” verb little v° , the temporal head T° and the complementizer C° . In German, the CP does not show exactly the same configuration as in the Romance languages and English. This is mainly due to the fact that German is a verb-second language displaying a V2 word order in the main clause and a V-final word order in the embedded clause, whereas the Romance languages and English show a rather consistent SVO word order. This has two main consequences: in the main clause, the German inflected verb raises to C° , hence triggering the V2 word order. In the embedded clause on the other hand, it remains in T° because C° is occupied by the complementizer. Since German shows a head-final configuration, the inflected verb will appear at the end of the embedded clause. Moreover, it has been shown that the German participle does not move out of V° . This means that it remains in the lexical domain, not entering the functional (extended) projection of the verb [1] (p. 86).

- | | | | | | | | |
|----|---|------|----------|---------|-------|------------|-------|
| 5. | Ich | habe | das Buch | der | Marie | geschenkt. | |
| | I | have | the book | the.DAT | Mary | given | |
| | ‘I have given the book to Mary.’ | | | | | | |
| 6. | ... dass | ich | das Buch | der | Marie | geschenkt | habe. |
| | that | I | the book | the.DAT | Mary | given | have |
| | ‘... that I have given the book to Mary.’ | | | | | | |

The examples in (5) and (6) are schematized in the tree-structures in Figures 3 and 4. The trees are not meant to represent antisymmetric trees; we adopt the representation proposed by [1]. The discontinuing arch delimits the functional from the lexical domain of the verb.

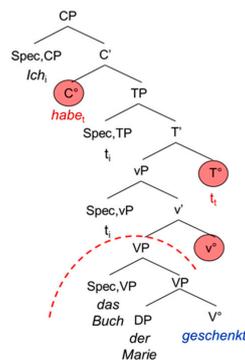


Figure 3. Structure of the German main clause.

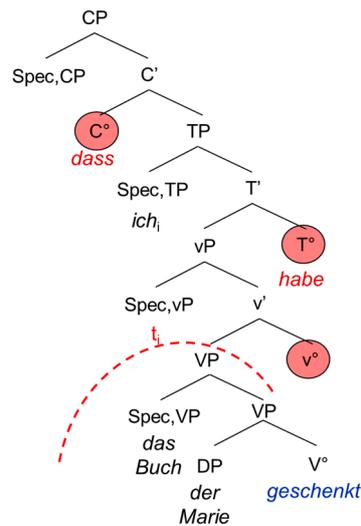


Figure 4. Structure of the German subordinate clause.

The French and English main and embedded clauses do not differ from each other apart from the complementizer that usually introduces the embedded clause: Both are head-initial with the inflected verb staying in T°. The participle moves to v°, hence entering the functional domain of the verb [1] (p. 95) (see also [20]). Again, examples (7) and (8) are represented by the tree-structures in Figures 5 and 6 below.

- 7. J' ai donné le livre à Marie.
I have given the book to Mary
- 8. ... que j' ai donné le livre à Marie.
... that I have given the book to Mary

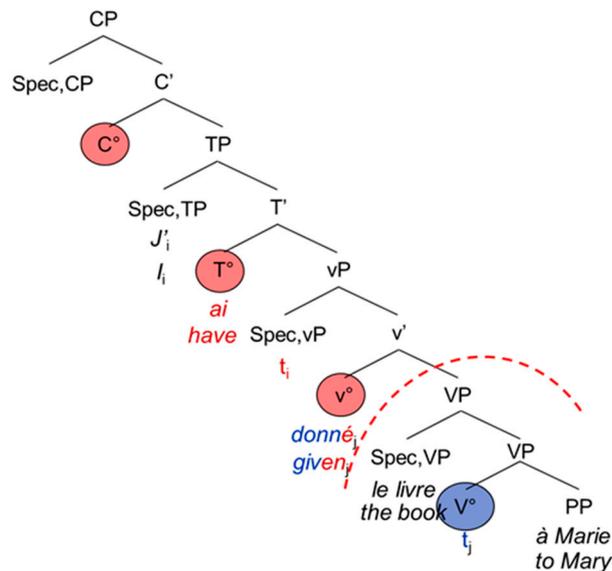


Figure 5. Structure of the French main clause.

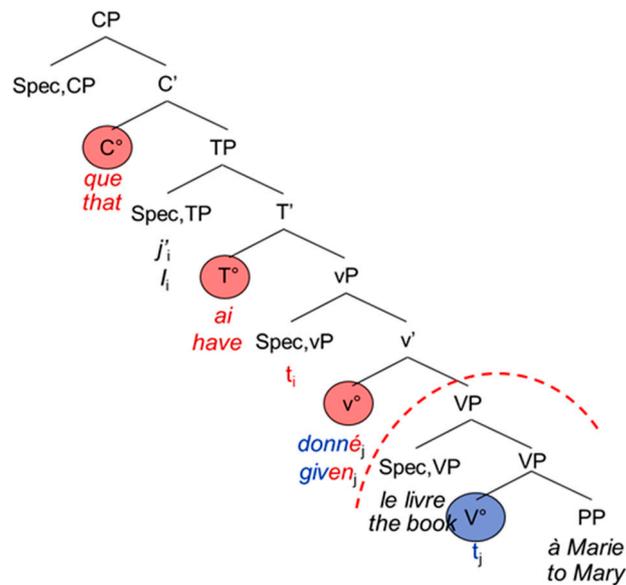


Figure 6. Structure of the French subordinate clause.

According to the PA, the theta filter (thematic roles, cf. [21]) must be satisfied and there is agreement (person, number) inside the CP (e.g., between the subject and the verb). The PFR further states that there should not be any CS in the functional domain, namely between the morphological exponents of the complementizer (C°) and the verbal morphology (the exponents of T° and v°). The differences between the syntactic architecture of German and that of French and English lead to interesting asymmetries in the possibility for CS, as [1] has been observed for German-Spanish bilinguals. Since the participle enters the functional domain in Spanish, but not in German, it is possible to switch the code for the German participle (9b), but not for the Spanish one (9a):

- 9. a. Er hat *contado.
- b. Ha erzählt.
 ‘(He) has told.’

Our grammaticality judgments as native German-French bilinguals reflect the same asymmetry. If we try to switch the code inside the periphrastic verbal construction of examples (6) and (8) above, we obtain the possible combinations in (10) of which only the alternative (10a) with the German lexical item *schenk-* endowed with the French past participle morphology *-é* is perfectly fine. Alternative (10b) does not seem completely impossible but is very marked; acceptability improves without a following indirect object (like in example (9b) above). This contrast is probably induced by the fact that the German participle, which remains in V°, follows all objects in the syntax of German (recall that it displays a V-final word order in embedded clauses). The two last alternatives with a French participle integrated in a German clause, both with French or German morphology are excluded.

- 10. a. ... que j'ai *schenké* le livre à Marie.
- b. ... que j'ai ??/**geschenkt* le livre à Marie.
- c. ... dass ich das Buch der Marie **donné* habe.
- d. ... dass ich das Buch der Marie **gedonnt* habe.

This pattern is expected under the PFR: we assume that lexical material merged in a low position can take part in syntactic operations (namely the German participle that can be endowed with French

morphology by moving to v°), while it is not possible to have a functional exponent (namely the French participle in v°) of a different language than the complementizer.

What sets González Vilbazo’s analysis [1] apart from the MLF, the MLF being arguably the most promising model for the last decade, is the claim that a ML is set on the level of extended projections instead of the whole CP. As a consequence, no language has to dominate since the phrase structure of every extended projection can be built according to a different grammar. González Vilbazo’s PFR differs also from the 4-M model (a refinement of the MLF, see Section 1.2) in the theoretical assumption about the “outsider late system morphemes” of the 4-M model [15,16]. Theoretically, the 4-M model predicts that structural case markers (i.e., for the nominative and accusative cases) should in fact originate from the same language as the verb’s functional morphemes, since these cases depend on the functional properties of the verb. According to the PFR on the other hand, it suffices that the features match. González Vilbazo himself does not discuss this point, but gives example (1), repeated as example (11), in which there is a German direct complement (*schlechte mündliche Noten* ‘bad oral marks’) paired with a Spanish inflected verb (*da* ‘gives’) [1]. Unfortunately, with plural nouns, the nominative and accusative cases are syncretic in German. However, if we replace the complement by a masculine singular noun in (12), the accusative morphology is apparent (*einen guten Kurs* ‘a good course’ vs. nom: *ein guter Kurs*).

- | | | | | | | |
|-----|--|---------------------|--------------------|-------------------------------|--------------------------------|--|
| 11. | El Lehrer,
the teacher
‘The teacher, which gives bad oral grades ... ’ | <i>que</i>
which | <i>da</i>
gives | schlechte
bad.F.PL.NOM/ACC | mündliche
oral.F.PL.NOM/ACC | Noten ...
grade.F.PL.NOM/ACC
[1] (p. 79) |
| 12. | El Lehrer,
the teacher
‘The teacher, which gives a good course ... ’ | <i>que</i>
which | <i>da</i>
gives | einen
a.M.SG.NOM | guten
good.M.SG.NOM | Kurs ...
course.M.SG.NOM
(our adaption) |

Moreover, in more recent publications, Myers-Scotton and Jake argue that complementizers are not all of the same kind, since their classification is based on a conceptually relevant versus system-relevant distinction [16] (pp. 351–354). Some complementizers are claimed to be late system morphemes (the “outsiders”) while others are content morphemes or both. The former usually appear in the ML (even if different from the CP they introduce) while the latter are more likely to occur in the language of the CP they introduce. As a consequence, and in contrast to González Vilbazo’s approach [1], the C° head could appear in a different language than the functional elements of the CP it introduces. But again, since it is not clearly defined which complementizers are outsiders and because there seems to be variation according to the languages involved in CS, this prediction is difficult to prove or falsify. However, if there is indeed a case of mismatch between a complementizer and the following functional material, this could falsify the PFR, but only if the connector is really a head. This last restriction is fundamental, as González Vilbazo points out [1] (pp. 81–84). For example, connectors conveying the same meaning in two languages are not automatically C-heads in both of them. There exist also different kinds of connectors bearing the same meaning in one and the same language, as is shown by the complementizer *weil* contrasting with the coordinating connector *denn*, both meaning ‘because, since’ in German, but only the first being a subordinating device, as can easily be seen by the German word order.

Having said that, Myers-Scotton and Jake also give some examples which are challenging for González Vilbazo’s claims [1], since they show that bridge connectors like English *that* vary according to whether they are morphological exponents of the language of the matrix CP, such as in (13) from a Palestinian Arabic–English CS, or of the embedded CP [16] (pp. 352–353).

- | | | | | |
|-----|---|--------------------------------|------------------------------------|--|
| 13. | [...] kaan
[...] PERF.3MASC.be
‘[he] was, the doctor, doubting that it was not reliable.’ | <i>el-doctor</i>
the-doctor | <i>yišuk</i>
IMPERF.3MASC.doubt | <i>it is not reliable</i>
<i>it is not reliable</i>
[22] (p. 71), in [16] (p. 353) |
|-----|---|--------------------------------|------------------------------------|--|

In the former case, Myers-Scotton and Jake propose to analyze the “entire multi-clausal constituent in one language” [16] (p. 353), i.e., the language of the matrix CP, and the following TP-clause is an

embedded island, and not the ML of the embedded CP. A CS from a complementizer to its following TP is unexpected under the PFR and has been argued several times (cf. [12], (pp. 224–225) among others) to be ungrammatical. Interestingly, the examples given in [16] are all with saying verbs like *say*, *confirm* or *doubt*. Now, it is well known that the bridge following this type of verb is a rather loose connector, introducing sentences that are less integrated in the matrix sentence. According to Haegeman [23], the relation between the matrix sentence and these loosely embedded sentences is more one of coordination than of subordination, which could explain the unexpected CS here: coordinating linking words are not subject to the PFR.

It is also worth considering that in the reported examples of CS between English and Arabic in [16], the Arabic complementizers equivalent of *that* show morphological agreement. We cannot discuss this issue in detail here, but it could point towards a difference in CS depending on whether the subordinating head expresses the Force-head or the Fin-head in Rizzi's split-CP hypothesis [24]. Rizzi [24] (p. 312) and Rizzi and Shlonsky [25] argue that only the Fin-head, but not the Force-head, can be associated with agreement. If the complementizers which do not match the language of the following TP are instances of Fin° , we could speculate that it is Force° which sets the ML of the CP and Fin° might not be bound to the PFR since it carries more lexical features (besides functional ones). If there can be shown such a correlation across languages, this would be welcome, otherwise, no principle can be stated for the complementizer's language in CS. However, these considerations have to remain speculative for the time being.

1.4. Code-Switching in Texting

The formal analysis discussed in this paper aims to complement and build on previous work pertaining to CS in text messaging, which now comprises a rather large number of studies on different contexts (for a more complete literature review see [8,26,27]; for research on texting in general see [28]).

A first body of research focuses on CS of individuals who regularly use different languages in all their everyday activities, including text messages, in language combinations like Tagalog–English in the Philippines [29], Spanish–English in the U.S. [30], English–isiXhosa in South Africa [31], English–Arabic in Kuwait [32], Nigerian English–standard English–vernaculars in Niger [33] and French–Wolof (and other languages) in Senegal [34]. Particular attention has been given to the influence of CS on the length of messages [29], assuming that CS leads to shorter messages, although this hypothesis has since been falsified [30,31]. These studies illustrate diversified and culturally specific CS practices and interpret them on a sociolinguistic background, assuming 'sociosymbolic values' of certain languages. However, they generally disregard formal aspects of CS and do not explicitly refer to any theoretical CS framework.⁶

A second body of research relates to work carried out within the project "sms4science", especially in Belgium [36], La Réunion [37] and Switzerland; in the Swiss context, CS was analyzed in text messages with Swiss German [26], French [8,27], Italian [38] and Romansh [39] as main text language (see also [7,40] for transversal and comparative elements).

These contributions use linguistic categories to describe the data mainly in a variational [8,26,40] and/or interactional [8,27] approach to language contact phenomena, but pay less attention to syntactic constraints on CS. Based on Poplack [11], the authors distinguish between intersentential and intrasentential CS. For the latter they further use Muysken's CS typology [41] to discriminate the insertion of elements originating from a language x into the utterance (or proposition) whose base language is y (called "transfer" by Auer [42] and "embedding" by Myers-Scotton [43]) and the alternation between two distinct base languages.

⁶ See however [35] who worked on Arabic-English CS in Jordanian texting and show that certain linguistic units are more used than others: they attest a high proportion of single nouns (34%), but also clauses (12%), phrases (21%) and functional categories such as conjunctions (8%), articles (7%), and pronouns (5%).

The results from analyses of language-specific subcorpora in the context of the Swiss *sms4science* project show that even though CS is quite frequent (57% of the Romansh messages, 23% of the Italian messages, 22% of the Swiss German messages and 18% of the French messages contain at least one switch), the patterns are minimal in nature, as most of the elements are very short, potentially formulaic intersentential alternations (e.g., *hola* 'hi', *ti amo* 'I love you', *besos* 'kisses', *what else?*).

Overall, more complex intrasentential alternations, intersentential CS and language change from one message to another are mainly found with older participants who consider themselves to be plurilingual.⁷ Due to the high proportion of short and recurrent intersentential alternations, often combined with recurrent language play, the usefulness of syntactic analyses of CS has been questioned for the corpus at hand (e.g., [8]) and syntactic constraints were not tested on any subpart of the corpus so far. However, in spite of the fact that the sequences of CS with which it is dealt are rather short and relevant cases perhaps limited in number, this does not imply that they could not be used to test hypotheses, especially if these focus on the transition point from one variety to the other in CS, since syntactic principles should apply to every kind of CS, regardless of their length or communicative intention.

Apart from that, a corpus of Swiss text messages seems to be an ideal ground for testing hypotheses on CS for two reasons. Firstly, the informal context, in which most of the text messages are written, facilitates CS with respect to other more formal contexts where CS is stigmatized. Secondly, in Switzerland, people not only switch from their mother tongue to English (which is taught as a foreign language at school)⁸, but also to other languages spoken (and regionally taught as L2) in the country, namely (Swiss) German, French, Italian and Romansh (see Section 2 for further details). This particular sociolinguistic situation allows us to test if the syntactic restrictions under study (the PA and the PFR, see Section 1.3) hold across different types of (informal) communication and different language pairs. If this is the case, it may be a strong hint towards the interlinguistic, if not universal validity of the principles.

As mentioned above, we are going to test the two proposed principles on a corpus of multilingual Swiss text messages. In the next section, we introduce the corpus and the methodology we used; the results are presented and discussed in Section 3.

2. Materials and Methods

Our corpus consists of 25,947 text messages, collected in Switzerland between 2010 and 2011. The corpus was compiled in the context of the project *sms4science.ch* [44]. In order to get genuine text messages, a publicity campaign was launched in order to invite people to send a copy of their original messages to a free phone number. In addition, the participants were sent a link to fill in an online form to provide sociodemographic information. The data were then anonymized and linked to the questionnaires by a code number.

More than 10,700 of the collected messages are written in a German dialect, about 7300 in near standard German, circa 4600 are in French, about 1500 show an Italian variety, and 1100 a Romansh one. The corpus also contains nearly 500 English messages as well as about 200 messages in other languages and varieties. These indications rely on the main language of the message, since, as mentioned before, about a quarter of them contain at least one CS element.⁹ The text messages were sent by 2627 different mobile devices. Note that in 2010/2011 when the corpus was built, smartphones were not widespread

⁷ For the Swiss German corpus, Bucher [26] also shows an effect of age, standard German alternations being used more often by older participants, English insertions being preferred by the younger ones (age is used as a continuous variable by Bucher). No analysis linked to L1 and L2 was carried out.

⁸ Following Cathomas et al. [40], English is used in the following proportions for the respective subcorpora: 44.1% for Swiss German, 59.1% for French, 28.7% for Italian, and 22.6% for Romansh.

⁹ For each text message, a main language was defined as the dominant language, i.e. the most used language in the message in terms of words (tokens).

in Switzerland, hence almost all messages were texted with a traditional mobile phone keyboard and, according to the questionnaires, without a spell checker.

The data were complemented by a manually standardized layer, which means that sequences were transcribed into standard orthography or, in the case of German dialects, translated into Standard German. Each token from a different variety than the main language of the message was manually attributed a corresponding tag.¹⁰

Subsequently, the subcorpora were parsed by an automatic Part-of-Speech (PoS) tagger, given the large amount of data. For this task, different taggers were used, since the available taggers are all trained for different languages. This fact leads to the inconvenient result that the subcorpora are tagged with different labels in different subcorpora. In addition, the accuracy of the PoS-tagging is slightly degraded because text message data display non-standard language, whereas the taggers are trained on near standard corpora. Specifically, repetitive code-switches induced the automatic taggers in error. These drawbacks have to be kept in mind for the interpretation of the results. However, in analyzing specific examples, the original messages were looked at and reported here in order to make sure that the analysis is not biased by standardizing issues.

For the present study, the data was searched semi-automatically using the search engine ANNIS (<http://corpus-tools.org/annis/>)¹¹, combining language labeling and PoS-tagging. Unfortunately, the described methodology did not allow for a quantification of the results because there are too many false positives and false negatives in the search results. Hence, we will offer a qualitative analysis showing the tendencies, but we make no claim to completeness. This is also the reason why we are going to discuss specific examples rather than provide statistical data; the quantitative analysis of the data is left to future research.

3. Results

In this section we report first general observations on CS in the corpus (Section 3.1). Then, we give some examples on which the PA (Section 3.2) and the PFR (Section 3.3) can be confirmed. Section 3.4 shows instances of CS that contradict Poplack's Equivalence Constraint [11] and Belazi et al.'s Functional Head Constraint [12]. In Section 3.5 we analyze apparent counterexamples to the PFR; we propose an alternative explanation for the unexpected data and discuss them with regards to other CS models.

3.1. General Observations

From earlier studies on the same corpus, we know that most of the code-switches are in fact isolated items (e.g., greetings, excuses, wishes etc.), insertions of a single item (cf. (14)), or abbreviations (e.g., *lol* 'laughing out loud', *cya* 'see you', *tottb* (*ti voglio tanto bene*) 'I wish you all the best') [40].

14. je viens pas à Neuch today!
I come not to Neuchâtel today
'I will not come to Neuchâtel today!'

More complex code-switches (of more than three tokens) are less frequent and the switch is often intersentential (even indicated by a full stop in (15)):¹²

¹⁰ A token is technically defined here as a single character or a series of characters delimited by a space on both its extremities.

¹¹ The sms4science.ch corpus is publicly accessible online via the corpus website (www.sms4science.ch). The data can be consulted using two different browsers, the *Korpusnavigator* and ANNIS2. As the latter allows for more specific searches (exploiting PoS-tagging, language tagging and the normalized language level), it was preferred to the first.

¹² E.g., in the French subcorpus, only 0.8% of the CS are intrasentential [8].

15. Grr anche noi ti amiamo tanto tanto mio geologo papa.
 grr also us.NOM you.ACC love.1PL so.much so.much my geologist daddy
Pourrais tu ramener tous les pijama rangé ds l armoire à Maria?
 Could.2SG you bring all the pijamas stored in the wardrobe of Maria?
 ‘Grr, we too like you so much my daddy geologist. Could you bring all the pijamas stored in
 Maria’s wardrobe?’

From a sociolinguistic point of view, the French subcorpus shows that the code-switch consisting of only one token are mostly English items and that they are more frequently used by people who are 25 years old or younger. Not surprisingly, the more complex code-switches are performed by people who declared to be plurilingual. With these people, the switched items are more often of a variety spoken in Switzerland (e.g., French, Swiss German or Italian) than English [8].

On the whole, from our research questions we can make three main observations which will be discussed in further detail in the following sections:

- We did not find any counterexample to the PA: where the agreement features are spelt out, they agree inside the DP and the CP.
- Overall, there are very few examples on which the PFR can be verified (or falsified). This is mainly due to the aforementioned fact that the vast majority of CS is intersentential.
- If there are switches inside an extended projection, they almost always occur on lexical heads. This is expected by the PFR. The few cases that do not conform to the PFR will be discussed in Section 3.5.

In what follows we will illustrate these observations with a detailed examination of some examples.

3.2. On Confirming the Principle of Agreement

In (16) below,¹³ there are two instances of CS, one for the greeting and one for the farewell:

16. *Ciaobella,* ischgestabiguetgange?
 hi.beauty, is.yesterday.evening.well.gone?
 händallwidrdierichtigijagemitheigno?
 have.all.again.the.right.jacket.with-them.home.taken?
 Heiichwünscheeuerholsamiundsunnerichitägufkreta . . . dickebacino.
 hey.I.wish.youPL.relaxing.and.sunflooded.days.on.Crete . . . bigM.kissM.DIM
 ‘Hi [my] beauty, has it gone well yesterday evening? Has everybody taken home the right jacket?
 Hey, I wish you relaxing and sunny days in Crete . . . Big kiss.’

The farewell is interesting, since it shows masculine agreement on the Swiss German adjective *dicke* ‘big’ according to the gender of the Italian diminutive *bacino* ‘little kiss’. The diminutives in Swiss German on the other hand are all neuter (*Küssli*, ‘little kiss’) and would trigger the neuter agreement morpheme *-s* on the adjective (*dicks*, ‘big’). This shows that the Principle of Agreement indeed applies: irrespective of the variety from which the morphemes originate, the noun has an intrinsic gender that triggers agreement on the adjective.

In (17) we have a similar case: the Italian proper noun *Ticino* triggers masculine agreement not only on the Italian adjective *bello* ‘lovely, beautiful’, but also on the Swiss German determiner *dä* ‘the’. The German corresponding *Tessin* ‘Ticino’ however is neuter, which would trigger neuter agreement on the determiner (*d)s* ‘the’:

¹³ We kept the original orthography of the messages. In this case, the writer chose not to separate the words by blank spaces).

17. Schön, dass äs dir guät goth & du dä *bello*
 nice that it you.DAT well goes & you the.M.SG lovely. M.SG
Ticino so chasch gnüssä
 Ticino. M.SG so can.2SG delight
 'Nice, that you are doing well and that you can so appreciate the lovely Ticino.'

Example (18) shows that even the English nouns like *shower* trigger, in this case, feminine agreement on the adjective, since it is most probably interpreted as inherently feminine as is the French *douche*:

18. ouf, ben quel timing! moi j ai pas arrêté de courir entre
 ouf well what.a timing me I have not ceased of run. INF between
 lessives, courses, ménage et théâtre. bises et bonne *baby shower!!!*
 laundries shopping housework and theatre kisses and good.F baby shower

Since Poplack et al. [9], it is consented that case assignment to borrowed or switched nouns depends on different factors. However, in the present case, there is no biological gender involved, nor does the noun's morphological pattern correlate with a specific grammatical gender (as it rarely does in French in general). Having excluded these clues, we are left with a high probability that gender assignment here is in fact due to analogy with the French equivalent, a factor that has turned out to be a strong indicator especially in languages which show little correlation between noun morphology and gender.

Of course, these are only a few examples in which the agreement pattern can be shown to hold across different varieties. The challenge here is to track morphologically visible agreement that is triggered by a noun (or a pronoun) which does not bear the same inherent case or number in both the varieties in order to verify the PA. In our corpus, this configuration appears either with Romance nouns whose German counterpart would bear neuter gender and an agreeing German adjective or article (cf. (16) and (17)), or with English nouns, which do not show overt gender at all with an agreeing German or Romance adjective or article (cf. (18)). But still, we did not come across any counterexamples to the PA in the corpus.

3.3. On Confirming the Principle of Functional Restriction

Turning to the PFR, consider (19):

19. J' ai pas looké mon *fone*
 I have not looked my phone
 'I did not look at my phone'

Here, a French auxiliary (*ai*) is accompanied by an English lexical verb (*look*), that is endowed with French participial morphology (*-é*). In Section 1.3 we have seen that this pattern is expected to occur: once the functional T^o-head is filled by the French auxiliary, the participle must also bear French morphology because it moves to the functional domain (i.e., to v^o). In contrast, the lexical item, introduced very low in the derivation (i.e., in V^o), can freely be of another variety than the functional heads. Tellingly, the verbal complement is linked to the verb according to the French subcategorization frame, namely as a DP, whereas the English frame would require the preposition *at*. This seems to confirm the principle, according to which the functional requirements must follow one and the same grammar.

In a similar vein, example (20) shows that even if German lexical material is introduced between the English auxiliary *do* and the participle, the latter preferably displays the English morphology:

20. how did *das neue jahr* begin?
 how did the new year begin
 'How did the new year begin?'

In principle, according to the PFR it would be possible to combine the *do*-auxiliary with a German infinitive (*beginnen* ‘to start’), as it remains in V°, even if standard German does not form the past tense with the auxiliary *do*.

21. [?]how did *das* *neue Jahr* *beginnen?*
 how did the new year begin.INF
 ‘How did the new year begin?’

However, it is completely banned to form a mixed structure that combines the German past perfect consisting of a *have*-auxiliary plus participle (22a) with the English *do*-auxiliary (22b). This is again predicted by the PFR since the functional T° head determines the shape of the following functional elements; the participle relating to an English T° head must rise to the functional domain and adopt English morphology.

22. a. *wie* *hat* *das* *neue Jahr* *begonnen?*
 how has the new year begin.PTCP
- b. **how* *did* *the* *new year* *begonnen?*
 how did the new year begin.PTCP

The PFR seems to hold even for further functional relations. For instance, in (23), the preposition *pour* ‘for’ governing the lexically English infinitive *ride* triggers French infinitival morphology (*-er*) on the verb. This is again in sharp contrast to the English counterpart, which takes a verbal *-ing* form after the preposition *for*, corresponding to the French *pour*.

23. *je* *dois* *aller* *m* *acheter* *une* *new* *veste* *pour* *rider* *ok?*
 I must go.INF REFL buy.INF a new jacket for ride.INF ok?
 ‘I have to buy myself a new jacket for riding, ok?’

But, as we shall see in the next section, the fact that a specific structure does not exist in one of the two CS-languages is not *per se* a prohibiting factor for switching.

3.4. Against an Equivalence Constraint and the Functional Head Constraint

As already mentioned in Section 1.2, the Equivalence Constraint proposed by Poplack [11] cannot explain the restrictions to which the syntax of CS obeys. In particular, we find switched lexical material in structures that only exist in one of the involved languages. For example in (24), the English tensed verb *have* takes a Swiss German complement that cannot be translated word by word into English because there is no corresponding phrasal verb to the standard or Swiss German *gern haben*, viz. *gärn ha* ‘to like’ (cf. (24b)).¹⁴

24. a. I haveyou *eifach* *molto* *gärn*
 I haveyou simply very -
- b. Ich han dich eifach sehr gärn!
 ‘I like you simply very much!’

Nevertheless, as we can see, the complement of the English verb consists of German lexical material. This is possible because the heads of the functional extended projection of the verb are all filled with English material. Moreover, the Swiss German complement *eifach* . . . *gärn* (‘simply . . . -’) meets the subcategorization requirements of the (English) verb. Hence, there is no restriction that would prohibit CS here.

¹⁴ This echoes the use of linguistic creativity in CS as addressed by Lakshmanan et al. especially for contexts where CS “is inclusive of hybrid structure and/or patterns that evince innovation and language change” [45] (p. 1). For recent sociolinguistic work arguing in favor of a shift away from additive models of bilingualism towards more holistic perspectives on different linguistic resources see, for instance [46]. In the specific context of texting see also [8].

Strictly speaking, according to the Equivalence Constraint [11], each occurrence of a composed verb form of a Romance language or English should not be compatible with a German complement; as we have seen in Section 1.3, German participles remain in a lower syntactic position than e.g., French or English participles and hence are placed at the end of a proposition, while in the other languages it precedes the complement. There is a (surface) ordering mismatch between the two types of languages involved and switching should be banned. But this prediction is not fulfilled, as we can see in (25). In (25a) we give the original message and in (25b) and (25c) we provide a French and a Swiss German translation, respectively. Not only does the Swiss German prefer the construction with the preposition *mit* ‘with’ like in English, but more crucially the complement precedes the participle, and this would also be the case if the complement was direct (i.e., without a preposition) in German.

25. a. Je t' appelle quand j' ai fini le *schwizertütsch.*
 I you.ACC call when I have finished the Swiss.German
- b. Je t' appelle quand j' ai fini le suisse allemand
 I you.ACC call when I have finished the Swiss German
- c. Ich lüte dir a wänn ich mit em Schwizertütsch fertig bin
 I call you VP¹⁵ when I with the.M.SG.DAT Swiss.German finished am
 ‘I call you when I’m done with the Swiss German [course].’

We have already mentioned that Belazi et al.’s [12] Functional Head Constraint (see Section 1.2) is not respected in our data. First, assuming that the determiner is a D°-head, it should only be able to select nouns of the same language, which is clearly not true as we have just seen in example (25). Further, example (20) above shows that the complement of the verb has not necessarily to be in the same language as the verb itself. Of course, the lexical verb V° is not a functional head. However, if we assume that the French infinitival and participial morphology stems from v°, then cases in which a foreign lexical verb is taking a French morphology should be banned, since the lexical verb V° is the complement of the functional verb v°. Nevertheless, example (19) above shows the contrary: the English lexical verb *look* is the complement of the functional verbal head v°. And this v° clearly shows French properties providing the verb with the French participial morpheme *-é*.

As for the 4-M model, its predictions are quite similar to the PA and the PFR. An empirical fact that could discriminate between the two models is the case marking on nominative or accusative constituents that would not correspond to the functional morphemes of the verb, as already discussed in Section 1.3. Our corpus provides rather limited exploitable data. This is mainly due to the fact that very often, the verbal morphology in the two languages involved in the switch is either homophonous as in example (26), or the chosen orthography can be interpreted as reflecting either the pronunciations, such as in example (27). In (26), the ending *-e* corresponds to both the English and the French inflection for the first person. In (27), the orthography <phone> can either be read as the English [fəʊn] or as a morphologized Swiss German [fəʊnə] with a schwa for the infinitive.

26. Je te *phone* demain pour t' engueuler! :)
 I you phone tomorrow for you give a roasting
 ‘I phone you tomorrow to give you a roasting.’
27. Wa söl i am Timo ez Säqe? CHasH mia *pHone*?
 what shall I to.the.DAT Timo now say Can.you me. DAT phone
 ‘What shall I say to Timo now? Can you phone me?’

Now, the 4-M model would only be contradicted if in these examples the English variant was intended: the agreeing morpheme on the verb would not correspond to the language of the case marked pronouns. But the fact that in (27) the complement pronoun *mia* is marked for dative hints

¹⁵ VP: verb particle.

towards a German like *phone*: while *phone* takes a direct complement in English, the Swiss German translation *alüte* takes an indirect complement.

3.5. How to Explain Apparent Counter-Examples to the Principle of the Functional Restriction

So far we have presented data that entirely support the two principles we are testing. However, we also found some data that appear problematic at first sight. Consider (28):

28. A proposito, sempre di quel capitolo, le trois grandes routes
 By the way, always of this chapter the.f.pl. three big.f.pl. roads
de la persuasion sono le 3 implicazioni (*forte, moyenne, nulle*)?
 of the persuasion are the 3 implications strong medium null
 'By the way, concerning the same chapter, the *trois grandes routes de la persuasion* are the 3 implications (strong, medium, zero)?'

Unexpectedly, this example shows an Italian determiner followed by a French quantifier. Relying on the extended projection of the noun as presented in González Vilbazo [1], the determiner and the quantifier are both functional heads staying in D° and Q°, respectively. These two heads belong to the same extended projection, so they should be filled by lexical exponents of the same variety.

In the present case, however, we know that the texter is a native Italian speaker. Our explanation for the theoretically impossible CS inside the nominal functional extended projection is based on the assumption that the author does not analyse the components of the nominal expression one by one; she rather reanalyses the sequence *trois grandes routes de la persuasion* 'three big roads of the persuasion' as one and the same proper noun relating to a specific philosophic or psychologic concept. This assumption is plausible, firstly because the writer is not French speaking and secondly because it seems to be a quotation of a concept that both the writer and the addressee have been taught in a lecture they have attended together. If our assumption is correct, the PFR is not any longer violated: the Italian determiner simply combines with a French noun. According to the MFL, the sequence would be analysed as an EL island. A further option to analyse this example is to regard numerals as lexical heads instead of quantifiers. Then they do not constitute a problem for the proposed principle at all, since lexical heads are not affected by the predictions. But assuming that numerals are lexical heads weakens the predictive force of the PFR, since there remain very few cases in which it actually can be violated.

Another explanation for (28) could rely on the phonetic similarity between the Italian and the French determiner (*le* [le] and *les* [le] respectively): even if the determiner is spelt the Italian way, it could also be interpreted as the French [le], in the sense of phonetic spelling (cf. [47]). Perhaps this second hypothesis is farfetched for (28), since this would be the only string that is not spelt according to the standard orthography in this example.

However, in (29) below, the "phonetic spelling" explanation is much more verisimilar. The message is obviously a conscious wordplay, mixing English and Swiss German:

29. maybe *wänn* you are *zrugg* in züri at that *uhr*time you
 maybe when you are back in Zurich at that clock.time you
 can just drop by for a *kaffee* if you like
 can just drop by for a coffee if you like
 'Maybe when you are back in Zurich at that time, you can just drop by for a coffee if you like.'

In theory, it should not be possible to combine the Swiss German complementizer *wänn* 'when' with the English tensed verb *are*. But still, we do not think that the PFR was violated. The lexical item *wänn* is ambiguous, especially if we take into consideration that Swiss German people tend to pronounce the English when [wen] as [vɛn], which in turn gives rise to the suspicion that *wänn* can actually be interpreted as both the Swiss German [væn] and the pseudo-English [vɛn].

In addition, a genuine (Swiss) German *wänn* would trigger the V-final position of the tensed verb, cf. (30a) vs. (30b). But the V-final position is completely out with an English inflected verb, cf. (30c).

30. a. when you [are_v back in Zurich] at that time
 b. wänn du zu dere Ziit [z'Züri zrugg bisch_v]
 c. *wänn you at that uhrtime [z'Züri zrugg are_v]

We take this as evidence that *wänn* in (28) cannot be the ordinary Swiss German complementizer but rather a spelling variant of the English complementizer. We want to emphasize that both the FHC and the MLF model also fail to give an alternative explanation faced with this example; here the predictions of all the approaches we have introduced in Sections 1.2 and 1.3 make the same predictions. Furthermore, according to Myers-Scotton and Jake [16] (pp. 352–354), adverbial-like complementizers are often qualified as content morphemes in the 4M-model. These tend to occur in the EL-language of CP2 (i.e., the embedded clause), which in this case is clearly English. In addition, it is not suitable to analyse the Swiss German complementizer *wänn* as an EL island, because the C° head is supposed to be in the matrix language of either the matrix CP or the embedded CP. However, the quantitatively predominance of English in both the sentences, as well as the English word order in the embedded clause do not indicate that one of the ML would in fact be Swiss German. It appears then that there is no need of the concept of EL islands here, since we are able to give an alternative, more insightful explanation in terms of language system for this example.

Finding a reasonable explanation might be more challenging with regards to the three examples of our corpus in which there seems to be a clash or rather a mixture of two distinct grammars, namely French and German grammar. First, consider (31):

31. Commenta gehts à zuri? Je viens te voir en janvier, oki?
 how it goes.it at Zurich I come you.ACC see in January okay?
 'How is it going in Zurich? I will come to visit you in January, okay?'

The sender mixes the two question tags meaning 'how are you?' of French *comment ça va?* and German *wie geht's?*, respectively. The problem is that they are incompatible in the sense that the inflected verb stays in different positions in the two grammars; as we have seen in Section 1, the German V2-condition triggers its movement to C° whereas in French it remains in T°. This leads to the fact that in the linear order the expletive subject of the question, *ça* and 's respectively, appear once before and once after the verb (even if both stay in Spec,TP, i.e., the position reserved for the subject). When mixing up the two languages in a linear order like in (31), the subject appears twice. This is ungrammatical in both languages. Therefore, for this example we want to propose an analysis that does not involve syntactic CS, but rather the insertion of a reanalyzed verbal item into the French syntax.

It is unlikely that the German verb *geht* moves to C° in the presence of the French preverbal subject *ça*, if we do not want to assume a sort of loop by which the computational process could run twice through one and the same syntactic position. Hence, we propose to analyze the whole German item *gehts* resulting from the contraction of *geht es* as the inflected verb. More precisely, the speaker inserts *gehts* like an infinitival form, meaning 'to go' in V° in the syntactic derivation. From there, the verb moves to v° and T° to acquire functional information. The verb has to agree with the third person singular in French grammar. Significantly, the most frequent and productive agreement morpheme (the morpheme for regular verbs) for third person singular is a zero morpheme in spoken French. If we are right and the item *gehts* is inserted as a chunk in the derivation, we expect it not to show overt agreement in French grammar (viz. *gehts+∅*).¹⁶

¹⁶ A reviewer points out that example (31) may be analyzed as a "blend representation" in the sense of Goldrick et al. [48]. The authors argue that a head, or alternatively an expletive constituent, can appear twice in a sentence with CS, i.e., once in each of the languages involved in CS. However, since their model is based on an optimality theory approach, high-ranking

This analysis clearly shows that the presence of an apparent syntactic CS (or grammatical CS) does not always imply an actual syntactic CS. Rather, grammar stops where unanalyzed (lexical) items are inserted in their actual form in the derivation, where in turn they are treated like any other item of the corresponding word class (e.g., verbs or nouns). These items, which are imitated but do apparently not carry the relevant syntactic features of the language they stem from, are probably better analyzed as real (syntactic) islands. This notion should however be restricted to CS containing syntactically not analyzed chunks. These can be integrated in, or modified by the grammar of the extended projection they belong to, whereas EL islands of the MLF model seem not to be affected by the ML at all.

4. Conclusions

In the present study we tested two syntactic principles on CS: the Principle of Agreement and the Principle of Functional Restriction. We have chosen to follow González Vilbazo [1] in considering the numerals and the determiners to be functional categories staying in the heads of the extended projection of D° and the verbal morphology to be the overt material of the functional categories of the extended projection of V° .

In our corpus of spontaneous non-standard written data we did not find convincing counterexamples to the principles. This is remarkable, because it is often claimed that in text messages everything is possible, but see [49–51] for evidence that syntactic restrictions also hold in this type of data. Even if we have a lot of CS in the multilingual corpus, they do not seem to be random: once a language (or a grammar) has been chosen for the syntactic computation of an extended projection, all the functional heads of this projection have to conform to the chosen language. In addition, the agreement and subcategorization requirements of the head of a projection are always satisfied, even if not obligatorily by morphologic exponents of the same language (especially when agreement is involved). The two principles seem to be robust, at least for CS between and inside Romance and Germanic varieties.

We have also seen that other proposed principles, namely Poplack's [11] Equivalence Constraint, Belazi et al.'s [12] Functional Head Constraint, or Myers-Scotton and Jake's 4-M model [15,16] make less precise predictions and some of these have been falsified in our corpus. Even if the 4-M model could stand up to the test with the data of our limited corpus, as discussed in Section 1.3, not all its predictions turn out to be tenable.

The fact that we could not find any single violation of the tested principles, which cannot be explained by other factors, is striking. While uncovering interesting tendencies, the present article has important limitations that need to be addressed. The data on which we tested the principles are scarce and we did not conduct a quantitative analysis. It must be stressed that violations may have been overlooked due to the semi-automatic searches on the corpus.

A second point relates to the language sample of which we disposed. Principles like the Equivalence Constraint may reveal themselves more useful when CS occurs between typologically more diversified languages. An investigation among Romance languages, English and German is perhaps too restricted in terms of variation and interplay that could reveal more insights into the core principles of CS. Nevertheless, we could identify some interesting interactions between a language of the V2 type (German) and the SVO languages in our corpus concerning word ordering issues. In sum, further investigation based on the (manual) annotation of a sample of more diversified languages and varieties may provide stronger evidence for universal syntactic restrictions on CS.

Acknowledgments: We would like to express our gratitude to the Swiss National Science Foundation (Berne, Switzerland) for financing the project "SMS communication in Switzerland: Facets of linguistic variation

constraints preclude expletives as the sole doubled element: "we should not observe doubling of expletive elements (e.g., English *do*) alone" [48] (p. 870).

in a multilingual country". Many thanks also to Elisabeth Stark (University of Zurich, Switzerland) for enabling this research project.

Author Contributions: A.R.-T. and E.M. conceived the research and explored the data together, A.R.-T. performed all the formal analyses of the examples and wrote most parts of the paper. E.M. contributed to the discussion, wrote parts of the introduction (1.1 and 1.2) and was involved in the formal revision of the paper.

Conflicts of Interest: The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

- González Vilbazo, K. Die Syntax des Code-Switching. Esplugisch: Sprachwechsel an der Deutschen Schule Barcelona. Ph D. Thesis, Universität zu Köln, Cologne, Germany, 2005.
- Vogt, H. Language contacts. *Word* **1954**, *10*, 365–374. [[CrossRef](#)]
- Gumperz, J.J. *Discourse Strategies*, 1st ed.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 1982.
- Auer, P. Bilinguales Sprechen: (Immer noch) Eine Herausforderung für die Linguistik. *Sociolinguistica* **2006**, *20*, 1–21. [[CrossRef](#)]
- Mondada, L. Le code-switching comme ressource pour l'organisation de la parole-en-interaction. *J. Lang. Contact* **2007**, *1*, 168–197. [[CrossRef](#)]
- Alvarez-Cáccamo, C. From 'switching code' to 'code-switching': Towards a reconceptualization of communicative codes. In *Code-Switching in Conversation: Language, Interaction and Identity*; Auer, P., Ed.; Routledge: London, UK; New York, NY, USA, 1998; pp. 29–50.
- Morel, E.; Bucher, C.; Pekarek Doehler, S.; Siebenhaar, B. SMS communication as plurilingual communication: Hybrid language use as a challenge for classical code-switching categories. *Linguist. Investig.* **2012**, *35*, 260–288.
- Morel, E. Le Bricolage Plurilingue Dans La Communication Par Texto: Interprétations D'une Pratique Entre Affiliation Locale et Aspiration Globale. Ph D. Thesis, Université de Neuchâtel, Neuchâtel, Switzerland, 2016.
- Poplack, S.; Sankoff, D. Borrowing: The synchrony of integration. *Linguistics* **1984**, *22*, 99–135. [[CrossRef](#)]
- Gardner-Chloros, P. *Code-Switching: An Introduction*, 1st ed.; Cambridge University Press: London, UK; New York, NY, USA, 2009.
- Poplack, S. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* **1980**, *18*, 581–618. [[CrossRef](#)]
- Belazi, H.M.; Rubin, E.J.; Toribio, A.J. Code switching and X-Bar Theory: The functional head constraint. *Linguist. Inq.* **1994**, *25*, 221–237.
- MacSwan, J. The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Biling. Lang. Cogn.* **2000**, *3*, 37–54. [[CrossRef](#)]
- Myers-Scotton, C. *Duelling Languages: Grammatical Structure in Codeswitching*; Clarendon Press: Oxford, UK, 1993.
- Myers-Scotton, C.; Jake, J. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* **2000**, *38*, 1053–1100. [[CrossRef](#)]
- Myers-Scotton, C.; Jake, J. A universal model of code-switching and bilingual language processing and production. In *The Cambridge Handbook of Linguistic Code-Switching*; Bullock, B., Toribio, A.J., Eds.; Cambridge University Press: London, UK; New York, NY, USA, 2009; pp. 336–357.
- Bhatt, C. *Die syntaktische Struktur der Nominalphrase im Deutschen*, 1st ed.; Gunter Narr: Tübingen, Germany, 1990.
- Alexiadou, A.; Haegeman, L.; Stavrou, M. *Noun Phrase in the Generative Perspective*, 1st ed.; Mouton de Gruyter: Berlin, Germany; New York, NY, USA, 2007.
- Cinque, G. *The Syntax of Adjectives: A Comparative Study*, 1st ed.; MIT Press: Cambridge, MA, USA, 2010.
- Cinque, G. *Adverbs and Functional Heads. A Cross-linguistic Perspective*; Oxford University Press: New York, NY, USA, 1999.
- Chomsky, N. *Lectures on Government and Binding: The Pisa Lectures*, 1st ed.; Foris Publications: Dordrecht, The Netherlands, 1981.

22. Okasha, M. Structural Constraints on Arabic/English Code-Switching: Two Generations. Ph D. Thesis, University of South Carolina, Columbia, SC, USA, 1999.
23. Haegeman, L. *Adverbial Clauses, Main Clause Phenomena, and the Composition of the Left Periphery*, 1st ed.; Oxford University Press: New York, NY, USA, 2012.
24. Rizzi, L. The fine structure of the left periphery. In *Elements of Grammar*; Haegeman, L., Ed.; Kluwer: Dordrecht, The Netherlands; Boston, MA, USA; London, UK, 1997; pp. 281–337.
25. Rizzi, L.; Shlonsky, U. Strategies of subject extraction. In *Interface + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*; Sauerland, U., Gärtner, H.-M., Eds.; de Gruyter: Berlin, Germany; New York, NY, USA, 2007; pp. 115–160.
26. Bucher, C. SMS-User als glocal Player: Formale und funktionale Eigenschaften von Code-switching in SMS-Kommunikation. *Networx* **2016**, 73.
27. Pekarek Doehler, S. Hallo! Voulez vous luncher avec moi hüt? Le “code switching” dans la communication par SMS. *Linguist. Online* **2011**, 48, 49–70.
28. Thurlow, C.; Poff, M. Text messaging. In *Pragmatics of Computer-Mediated Communication*; Herring, S.C., Stein, D., Virtanen, T., Eds.; Mouton de Gruyter: Berlin, Germany; Boston, MA, USA, 2013; pp. 163–188.
29. Bautista, M.L.S. Tagalog-English code switching as mode of discourse. *Asia Pac. Educ. Rev.* **2004**, 5, 226–233. [[CrossRef](#)]
30. Carrier, M.; Benitez, S. The effect of bilingualism on communication efficiency in text messages (SMS). *Multilingua* **2010**, 29, 167–183. [[CrossRef](#)]
31. Deumert, A.; Masinyana Oscar, S. Mobile languages choices: The use of English and isiXhosa in text messages (SMS). *Engl. Worldw. J. Var. Engl.* **2008**, 29, 117–147.
32. Haggan, M. Text messaging in Kuwait. Is the medium the message? *Multilingua* **2007**, 26, 427–449. [[CrossRef](#)]
33. Chiluba, I. Assessing the Nigerianness of SMS text-messages in English. *Engl. Today* **2008**, 24, 51–56. [[CrossRef](#)]
34. Vold Alexander, K. Analyzing multilingual texting in Senegal: An approach for the study of mixed-language SMS. In *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse*; Sebba, M., Mahootian, S., Jonsson, C., Eds.; Routledge: New York, NY, USA, 2012; pp. 146–169.
35. Al-Khatib, M.A.; Sabbah, E. Language Choice in Mobile Text Messages among Jordanian University Students. *SKY J. Linguist.* **2008**, 21, 37–65.
36. Cougnon, L.-A. *Langage et SMS: Une Étude Internationale Des Pratiques Actuelles. Cahiers du Cental. 8*, 1st ed.; Presses universitaires de Louvain: Louvain-la-Neuve, Belgium, 2015.
37. Ledegen, G.; Richard, M. “jv me prendre un bois monumental the wood of the century g di”: Langues en contact dans quatre corpus oraux et écrits “ordinaires” à la Réunion. *Glottopol* **2007**, 10, 86–100.
38. Moretti, B.; Stähli, A. L'italiano in contatto con il dialetto e altre lingue. Nuovi mezzi di comunicazione e nuove diglossie. *Linguist. Online* **2011**, 48, 71–82.
39. Grünert, M. Varietäten und Sprachkontakt in rätoromanischen SMS. *Linguist. Online* **2011**, 48, 83–113.
40. Cathomas, C.; Bucher, C.; Morel, E.; Ferretti, N. Same same but different: Code-Switching in Schweizer SMS—ein Vergleich zwischen vier Sprachen. *TRANEL* **2015**, 63, 171–189.
41. Muysken, P. *Bilingual Speech: A Typology of Code-Mixing*, 1st ed.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2000.
42. Auer, P. *Bilingual Conversation*, 1st ed.; Benjamins: Amsterdam, The Netherlands, 1984.
43. Myers-Scotton, C. Constructing the frame in intrasentential code-switching. *Multilingua* **1992**, 11, 101–127. [[CrossRef](#)]
44. Stark, E.; Ruef, B.; Ueberwasser, S. Swiss SMS Corpus (2009–2014). University of Zurich. Available online: <https://sms.linguistik.uzh.ch> (accessed on 10 November 2016).
45. Lakshmanan, U.; Balam, O.; Bhatia, T.K. Introducing the special issue: Mixed verbs and linguistic creativity in bi/multilingual communities. *Languages* **2016**, 1, 9. [[CrossRef](#)]
46. Otheguy, R.; García, O.; Reid, W. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Appl. Linguist. Rev.* **2015**, 6, 281–307. [[CrossRef](#)]
47. Anis, J. Neography: Unconventional Spelling in French SMS Text Messages. In *The Multilingual Internet. Language, Culture, and Communication Online*, 1st ed.; Herring, S., Ed.; Oxford University Press: Oxford, UK, 2007; pp. 87–115.

48. Goldrick, M.; Putnam, M.; Schwarz, L. Coactivation in bilingual grammars: A computational account of code mixing. *Biling. Lang. Cogn.* **2016**, *19*, 857–876. [[CrossRef](#)]
49. Dürscheid, Chr.; Stark, E. Anything goes? SMS, phonographisches Schreiben und Morphemkonstanz. In *Die Schnittstelle von Morphologie und geschriebener Sprache*; Neef, M., Scherer, C., Eds.; Mouton de Gruyter: Berlin, Germany, 2013; pp. 189–209.
50. Robert-Tissot, A. Le Sujet et son Absence dans les SMS Français. Une Analyse Syntaxique Basée sur le Corpus sms4science Suisse. Ph D. Thesis, Universität Zürich, Zürich, Switzerland, 2016.
51. Stark, E.; Robert-Tissot, A. Subject drop in Swiss French text messages. *Linguist. Var.* **2017**, in press.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).