Code switching, lexical borrowing, and polylanguaging in Valencian Spanish: an analysis of data from conversational corpora and Twitter

Andrew Jordan Lavender
University at Albany, State University of New York, jjlavender@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Anthropological Linguistics and Sociolinguistics Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1871

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
CODE SWITCHING, LEXICAL BORROWING, AND POLYLINGUISTIC IN
VALENCIAN SPANISH: AN ANALYSIS OF DATA FROM CONVERSATIONAL
CORPORA AND TWITTER

by

Andrew Jordan Lavender

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Languages, Literatures and Cultures
2017
CODE SWITCHING, LEXICAL BORROWING, AND POLYLANGUAGING IN VALENCIAN SPANISH: AN ANALYSIS OF DATA FROM CONVERSATIONAL CORPORA AND TWITTER

by

Andrew Jordan Lavender

COPYRIGHT 2017
ACKNOWLEDGEMENTS

First, I would like to express gratitude to my committee members: Professor Maurice Westmoreland, Professor Lotfi Sayahi, and Professor Cynthia Fox for their helpful comments and critique to complete this dissertation. I would like to thank Professor Westmoreland, in particular, for his guidance during the long process of writing this dissertation through our many private conversations.
ABSTRACT

This study examines lexical borrowing, code switching, and polylanguaging in Valencian Spanish to better understand how each is used differently in oral conversation in comparison with online communication on Twitter. This study compares data collected from three published corpora of oral interviews of speakers of Valencian Spanish with data collected from Twitter profiles of individuals residing in Valencia. In each of the sources Spanish is the preferred code into which Valencian material is inserted. A unique feature of data from the published corpora is the high frequency of code switching (CS) into Valencian in instances of reported speech. With regard to frequency, Twitter users switch from Spanish into Valencian, followed by from Valencian into Spanish and then from Spanish into English. On Twitter, the most frequent type of switch found is the tag switch, which includes exhortatives, greetings and farewells, happy birthday wishes, and a variety of other types of tags and other idiomatic expressions used in a highly emblematic fashion as a way of preforming identity. Both intrasentential and intersentential switches also appear online and reflect how discourse might be organized differently online than offline. In looking at lone vs. multiword insertions, the importance of turn taking is noted and instances where speakers are not in a naturalistic conversation evidence traits which influence patterns of CS and polylanguaging. Additionally, lexical economy is suggested as a motivating factor for CS on Twitter given the platform’s technological limitation of 140 characters per tweet.
## LIST OF TABLES

Table 1. Audience roles ........................................................................................................45
Table 2. Language use per individual user ..........................................................................69
Table 3. CS and Borrowing from Valencian into Valencian Spanish by constituent in PRESEEA
...........................................................................................................................................73
Table 4. CS and Borrowing from Valencian into Valencian Spanish by syntactic constituents in CSDCLP.................................................................................................78
Table 5. CS and linguistic borrowing from Valencian into Valencian Spanish per syntactic constituents in Val.Es.Co ..........................................................................................83
Table 6. Embedded languages on Twitter ...........................................................................86
Table 7. Tags, intersentential and intrasentential mixing on Twitter ..................................88
Table 8. Constituents ‘mixed’ on Twitter .............................................................................91
Table 9. Tags by function on Twitter ..................................................................................97
Table 10. Tags used on Twitter ...........................................................................................97
Table 11. Single vs. multiword insertions ..........................................................................111
Table 12. Number of informants with language contact phenomena per medium ..............117
Table 13. Most frequently used borrowed terms in the data sources ..................................122
Table 14. Valencian words in comparison with Castilian ..................................................124
Table 15. Orthographic economy in lone insertions on Twitter .......................................125
LIST OF FIGURES

Figure 1. Factors influencing language choice (Grosjean 1982: 136) .................................................. 13
Figure 2. Reasons for CS (Grosjean 1982) .......................................................................................... 24
Figure 3. Map of Valencia (province) .................................................................................................. 60
Figure 4. Map of Valencia and surrounding area (city) ........................................................................... 62
Figure 5. Reasons for CS (Grosjean 1982) .......................................................................................... 114
Figure 6. Reasons for CS updated for Twitter ...................................................................................... 115
Figure 7. Distribution of CS/language mixing among Twitter users ....................................................... 118
Figure 8. Distribution among informants in the PRESEEA corpus ....................................................... 119
Figure 9. Distribution of CS among informants in the CSDCLP ............................................................ 119
Figure 10. Distribution of CS among informants in Val.Es.Co .............................................................. 120
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................. iii
ABSTRACT .................................................................................. iv
LIST OF TABLES .......................................................................... v
LIST OF FIGURES ......................................................................... vi
TABLE OF CONTENTS ................................................................. vii

CHAPTER ONE ............................................................................. 1
INTRODUCTION ........................................................................... 1
  1.1 Introduction ........................................................................... 1
  1.2 Objectives and Research Questions ....................................... 1
  1.3 Linguistic background of the Valencia region ....................... 2
  1.4 Overview of the study ........................................................ 6
  1.5 Outline of chapters ........................................................... 7
  1.6 Conclusion ............................................................................ 9

CHAPTER TWO ............................................................................. 11
LANGUAGE CONTACT, CODE SWITCHING AND BORROWING ............ 11
  2.1 Language contact and bilingualism .................................... 11
  2.1.1 Language choice .......................................................... 12
  2.2 Borrowing ......................................................................... 14
  2.3 Code switching ............................................................... 18
    2.3.1 Linguistic factors in code Switching .............................. 18
    2.3.2 Extra-linguistic factors in code switching ....................... 23
  2.4 Code switching vs. borrowing ........................................... 27
  2.5 Conclusion ......................................................................... 30

POLYLANGUAGING, CMC, TWITTER ........................................... 32
  3.1 Translanguaging and polylanguaging ................................ 32
  3.2 Social media .................................................................... 35
  3.3 Multilingualism on social media sites ................................. 39
  3.4 Audience Design on Twitter and language use .................... 44
  3.5 Further considerations for CMC research ............................ 51
  3.6 Conclusion ........................................................................ 53

CHAPTER THREE ........................................................................ 54
METHODS .................................................................................. 54
  4.1 Introduction ....................................................................... 54
  4.2 Published conversational corpora ..................................... 54
    4.2.1 Corpus sociolingüístico de Castellón de la Plana (CSCDLP) .................................................................................. 55
    4.2.2 PRESEEA (Proyecto para el estudio sociolingüístico del español de España y de América) .................................................... 56
  4.2.3 Val.Es.Co ...................................................................... 58
  4.3 Twitter corpus .................................................................. 58
  4.4 Data analysis and coding ................................................... 64
RESULTS AND DISCUSSION

5.5.6.2 Nouns

5.5.6.1 Nonce borrowings

5.5.5.7 Polylanguaging into German, Portuguese, French and Italian

5.5.5.5 Motivations for Tag Switches and polylanguage

5.5.5.4 Tags with an explitive function

5.5.5.2 Greetings and farewells

5.5.5.1 Tags with exhortative function

5.5.5 Tag Switches and polylanguaging

5.5.4.3 Discourse markers

5.5.4.2 Nonce borrowed nouns

5.5.4.1 Established loanwords

5.5.4 Lexical borrowing in PRESEEA

5.5.3.1 Discourse markers

5.5.3. Nonce borrowings in the CSDCLP data

5.5.3.1 Nouns

5.5.3.2 Adjectives

5.5.3.1.1 Nouns

5.5.3.1.2 Adjectives

5.5.3.1.3 Adverb phrases

5.5.3.2.1 Reported speech

5.5.3.2.2 Intersentential code switching

5.5.3 Intersentential code switching

5.5.2.1 Nouns

5.5.2.3 Intersentential code switching

5.5.2.1.1 Nouns

5.5.2.3.1 Reported speech

5.5.2.3.2 Intersentential code switching

5.5.2 Intersentential code switching

5.5.1.1 Hashtags

5.5.1.2 Discourse markers

5.5.1.3 Reported speech

5.5.1.4 Intersentential switches

5.5.1 CSCDLP

5.5.1.1 Hashtags

5.5.1.2 Discourse markers

5.5.1.3 Reported speech

5.5.1.4 Intersentential switches

4.5 Conclusion

3.1 Introduction

2.4 Lexical borrowing in PRESEEA

2.4.1 Established loanwords

2.4.2 Nonce borrowed nouns

2.4.3 Discourse markers

2.3 Polylanguaging into German, Portuguese, French and Italian

2.3.1 Motivations for Tag Switches and polylanguage

2.3.2 Tags with an explitive function

2.3.2.1 Greetings and farewells

2.3.2.2 Tags with exhortative function

2.3.2 Tag Switches and polylanguage

2.2 Polylanguage switches

2.1 Introduction
CHAPTER ONE
INTRODUCTION

1.1 Introduction

The modern world has created a new sort of social reality through the ubiquity of social media in the lives of language users. Social media sites create a new form of digital “contact zone” (Canagajarah 2013: 30), in which users of different backgrounds interact with each other. Through an analysis of computer-mediated communication (CMC), linguists can analyze the effects of language contact. This study analyzes the contact phenomena of lexical borrowing, code switching (CS), and polylanguaging as they appear in offline and online sources in Valencian Spanish, a regional variety of Spanish influenced by, and in contact with Valencian, the western dialect of Catalan spoken also in and around Valencia, Spain. The phenomena of code switching and borrowing have long been an established focus of studies on linguistic contact; this study will look also at the phenomenon of polylanguaging, a third contact phenomenon which is often associated with language use on CMC where there is no assumption of user bilingual status or proficiency.

1.2 Objectives and Research Questions

This study analyzes lexical borrowing and CS as they appear in Valencian Spanish by examining the use of the two languages (the Valencian Spanish and Valencian) and the insertion of both multiword and lone items (either as instances of CS nonce borrowings or arguably as
established lexical borrowings) from either Valencian or English into Spanish. The study will address the following research questions:

1) In Valencian Spanish, what patterns exist of CS and lexical borrowing from Valencian in terms of constituent class of switched elements, of intrasentential, intersentential and tag switching, and of lone and multiword insertions?
2) In what ways do the three published corpora of oral interviews offer differing results regarding the contact phenomena mentioned above?
3) How do the results from the three published corpora of oral interviews differ from those found in the corpus of Twitter data?
4) What factors explain any noted differences between the four corpora (online and offline)?
5) To what extent are concepts associated with an analysis of contact phenomena of spoken speech applicable to an analysis of similar phenomena found in Twitter data?

1.3 Linguistic background of the Valencia region

Spanish and Catalan are two Romance languages spoken in Valencia. Valencian is the dialect of Catalan spoken in Valencia, and Catalan is spoken additionally principally in Catalonia and the Balearic Isles, with Central Catalan being the most standard form spoken in Barcelona. Valencian differs from Catalan in its phonology and some aspects of its morphology and lexicon and is more typologically similar to Spanish even more so than other varieties of Catalan with its similar phonology. Valencian has the additional palatal affricate and fricative phonemes: /ʒ/ and /ʃ/, which Spanish lacks. Additionally, it has the open vowels: /ɛ/ and /ɔ/, which further distinguish it from Spanish. However, it does not have the unstressed schwa as found in Catalan. In its morphology, it both diverges from Central Catalan and from Spanish in some key areas. The first person present tense generally ends in –e. The imperfect subjunctive differs significantly from Central Catalan with the following: que ell vinguera (that he come/que él viniera). It also resembles Spanish with regard to its preference for the simple preterite (cantà), rather than the periphrastic past (va cantar) used in Central Catalan.
Spanish and Catalan have co-existed for centuries in both Catalonia and Valencia. This prolonged contact has left traces on both languages and the study of the effects of that contact on both languages is extensive (Galindo Solé 2003; Gómez Molina 1997; Sinner and Wesch 2008; Wesch 1996, 2000; Vann 1995, 2001; among many others). Many studies have focused on the effect that Spanish has had over spoken Catalan (Boix 1993; Sanz and González 1995) and in written Catalan on the Internet (Stokes 2011), while others have focused on the effect of Catalan on the Spanish spoken in Catalonia (Casanovas 1996; Moll 1961; Romera 2003; Sinner 2004; Wesch 1996). Many studies have analyzed the effects of Valencian on the Spanish spoken there (Blas Arroyo 1992; Briz 2004; Gómez Molina 1997) where one finds features of a Valencian-influenced Spanish.

Valencian, a dialect of western Catalan, is a minority language both in Spain and in the region where it is spoken. Simpson (1981) mentions some general characteristics of minority languages. Some of these characteristics are not applicable to the linguistic situation in Valencia, however; it is important to analyze common traits of regional languages.

“It is not the language of all areas of activity indulged in by its speakers: for example, it may be excluded from administration or education, being confined to home, religious life or literature. It may live in the shadow of a culturally dominant language, dominant because of political, educational, social or religious factors… Bilingualism is a characteristic of its speakers… There may not be a recognized norm for communication in the minority language; that is to say, there may be no ‘standard language’” (235-236).

Valencia (province) is divided into 34 regions, whose use and knowledge of Valencian differs considerably. There are ten regions that are historically Spanish-speaking and are situated in the interior part of the region. Around 10% of the population lives in this Spanish-speaking region. The Valencian-speaking sub-regions are divided into five linguistic zones based on frequency of use of Valencian:
(1) Valencia (el Camp de Túria, el Camp de Morvedre, la Ribera Alta, la Ribera Baixa and la Costera)
(2) the central region of Alcoi-Gandia (Vall d’Albaida, la Safor, el Comtat, L’Alcoià and la Marina Alta)
(3) Castellón (Els Ports, Alt Maestrat, Baix Maestrat, Alcalatén, Plana Alta and Plana Baixa)
(4) Metropolitan area of Valencia (city, Horta Nord, Oest and Sud)
(5) Alicante (Marina Baixa, Alacantí, Baix Vinalopó, Vinalopó Mitjà)

The level of usage of Valencian differs in each of these linguistic zones and varies considerably depending on context. Recent data shows that the population of the Valencian Community is around 4.9 million residents, of those; approximately 2.4 million are native speakers of Valencian (Ethnologue). Recent surveys of Valencia, Castellón and Alicante (AVL 2005; GV 2010; SIES 2005) show that around 70% of the population understands Valencian while there is a lesser degree of the population capable of speaking, reading and writing in Valencian. Around 50% of the survey participants report to speak Valencian. The Generalitat Valenciana 2010 surveyed 6,600 individuals in Castellón, Valencia, and Alicante. The survey reports that many self-report to have a “limited perception” of Valencian (93.5%) and similar numbers for limited comprehension and basic writing skills. However, there has been a decrease in the use of Valencian in metropolitan areas such as in Valencia (city) and Alicante. However, this decrease is likely due to an influx of immigrants from other Spanish-speaking countries, and other parts of Spain, to Valencia and Alicante. Speakers prefer to use Valencian when the relationship with the interlocutor is personal and more intimate (AVL 2005; GV 2010; SIES 2005).

The linguistic situation in Valencia has often been interpreted as a case of extended diglossia and the data from recent surveys supports such an interpretation based off the preference for its use in certain contexts (Blas Arroyo 1997; Hudson-Edwards 1992). Valencian is regarded as the low language and Spanish being the language of prestige, although the situation is much more complex. Casenoves Ferrer (2011) conducted a matched guise study to
determine if a positive attitude towards Valencian would cause an increase in usage. The matched guise was performed in public and private secondary schools to evaluate students’ linguistic attitude towards the Valencian language. The results of her study show that Valencian is no longer associated with lesser-educated people but is now ranked as positive as Spanish, which has been the prestige language in the region, but this increase in prestige has not been accompanied by an increase in usage. There has been an overall decrease in its use in various domains from 1995 to 2005 (AVL 2005; GV 2010). However, there has been an increase in the skills of reading and writing in Valencian, and that the ability to comprehend, read, and write increased in the period from 1993-2006 (61), likely due to its introduction into the educational system (Armau 2013).

Recent studies have also been conducted to analyze the situation and vitality of Valencian in the region. Casenoves Ferrer (2010) conducted surveys with 320 students in Valencia (city) and Xàtiva identifying the level with which the students used Valencian and Spanish and with which language they identified. In her survey of Valencia (city), she notes that around 75% of the respondents are Spanish-dominant, while around 25% are primarily Valencian speaking preferring to speak Valencian with their parents and other family members. Her study also reports “language of identification” which does not correlate to language use. When this aspect is analyzed, the results show that the number of students who identify with Valencian is higher than those who speak it regularly. However, her results show that Valencian is still associated with familial contexts and, to a certain degree, with government functions but it is otherwise not associated with more public contexts. Other factors, such as adopting a ‘Catalan identity’ can influence language choice (Casenoves Ferrer 2011). Casenoves Ferrer (2011) suggests that various external factors might be significant in explaining language choice data, such as a change
in the composition of the population in Valencia brought about by the arrival of large numbers of immigrants from other parts of Spain.

The linguistic situation of Valencian is similar to what has been reported in other regions with the same relationship between a regional language and national language. O’Rourke (2005) studied the linguistic attitudes and practices of Irish and Galician students and noted that the students expressed positive attitudes and identity value to their respective regional languages but did not self-report an active use of either.

1.4 Overview of the study

This research examines how speakers of the variety of Valencian Spanish who have a range of different skill levels in Valencian use it as a linguistic resource in their conversational speech and in their social media CMC practices.

The study will first provide a quantitative analysis of the contact language phenomena of code switching (CS) and borrowing analyzed in three bilingual corpora containing spontaneous speech and in a corpus of Twitter data. The term “language mixing” is used (and explained in further detail below) to differentiate written CS found on Twitter from conversational CS found in the conversational corpora. Additionally, the phenomenon of polylanguaging (Jørgensen et al 2011) is analyzed particularly with regard to the Twitter data. The Twitter data will offer a way of exploring the use of words and structures taken from another language when it is likely that the informant is not proficient in that language in cases of polylanguaging. It will quantify the CS and borrowed items as well as compare intersentential and intrasentential switches. It will discuss how to determine each when analyzing online data. It will detail how tag switches relate to intersentential and intrasentential switches, as well as in coding established vs. nonce
borrowings. It will report the frequency of each type of CS and the absolute and relative frequencies of lone versus multiword insertions for each data source. Individual informants who have more instances of CS in their data sample will be compared to other informants to insure that their results do not skew the overall results of the study. The issue of economy will be explored as a possible motivation for including single word insertions on Twitter.

In order to better understand differences between CS and borrowing in two different mediums of communication, the following three corpora will be the source for conversational data as a basis for comparison with the Twitter data. (1) the Corpus sociolingüístico de Castellón de la Plana y su área metropolitana (CSCDP) (Blas Arroyo et. al. 2009), (2) the PRESEEA project in Valencia, El español hablado de Valencia. Materiales para su estudio (Gómez Molina 2001, 2005, 2007), and; (3) the Val.Es.Co. corpus (Valencia, Español, Coloquial) (Cabeledo and Pons 2011). By using three different sources for the conversational data, a fuller understanding can be arrived at on the patterns found in conversational speech.

1.5 Outline of chapters

The following is a brief outline of each of the following chapters and the topics included in each.

Chapters Two and Three detail the theoretical trends and findings in contact linguistics, lexical borrowing and CS, and the differences between CS, established vs. nonce borrowing, and polylanguaging. Chapter Two provides an overview of prior scholarship on CS and lexical borrowing as they have related primarily to the analysis of spoken speech. Chapter Three focuses on recent scholarship related to analyzing contact phenomena present in various forms of CMC. These chapters introduce the principal concerns and recurring questions on the nature of CS,
borrowing and polylinguaging. Chapter Three additionally provides an overview of scholarship which argues for a reformulation of many of the concepts introduced in Chapter Two when they are applied to the context of computer-mediated communication (CMC). Topics discussed in this regard derive from audience design theory, superdiversity studies, analyses of context-collapse, and other issues relating to multilingualism online.

Chapter Four describes the details of the corpora used and notes the strengths and weaknesses of each as a data source. It discusses issues related to coding established and nonce borrowings versus CS/language mixing and polylinguaging, and linguistic factors associated with each phenomenon. It lays out the techniques of analysis that will be applied across the different corpora. It situates this study within a framework of CMC research by comparing the methods used here to collect data with other studies on language mixing in CMC. Researchers of CMC have noted both the value and necessity of utilizing an online ethnography approach to collecting data, where the researcher closely observes the patterns of language use employed by a user.

Chapter Five offers the results from a quantitative analysis of the data, examining the occurrences of CS and borrowing in each data source, noting the syntactic constituents of each instance of CS and lexical borrowing as well as the frequency of intersentential switches, intrasentential switches, and tagged switches. The intent is to identify patterns for each in terms of source as well as the specific linguistic contexts, which favor different types of switching and borrowing. Language use was quantified on Twitter by noting the number of insertions into Spanish and Valencian as well as English. Other relevant phenomena such as the presence of tags vs hashtags were discussed in conjunction with existing research on CMC. The preference for single word and multiword insertions were compared in order to analyze the complexity of
linguistic material used in each insertion. The distribution of language phenomena among individual informants was quantified and analyzed to investigate what patterns differentiate high-frequency CS users from others. The use of English is analyzed as both a potential source for established borrowings and polylanguaging, and the issue of linguistic economy was examined as a potential linguistic motivation for CS on Twitter.

Chapter Six presents the answers to the different research questions and offers concluding observations based on the data analysis and a final discussion of relevant factors effecting CS, lexical borrowing, both established and nonce, and polylanguaging from the different data sets. The challenges and limitations of the study were noted as well as potentially beneficial avenues for future research.

1.6 Conclusion

This chapter has introduced the objectives of this study by presenting the research questions answered and an overview of the material found in each chapter. The language phenomena analyzed in this study are: lexical borrowing, both established and nonce, code switching, and polylanguaging in Valencian Spanish in both conversation and on Twitter, taking into consideration the different motivations for code switching in conversation versus its inclusion in written text. It also analyzes the insertion of both multiword and lone insertions from either Valencian or English into Spanish. Three conversational corpora were analyzed (1) the Corpus sociolingüístico de Castellón de la Plana y su área metropolitana (CSCDP) (Blas Arroyo et. al. 2009), (2) the PRESEEA project in Valencia, El español hablado de Valencia. Materiales para su estudio (Gómez Molina 2001, 2005, 2007), and; (3) the Val.Es.Co. corpus (Valencia, Español, Coloquial) (Cabebo and Pons 2011) in conjunction with a corpus of data collected from Valencian Twitter profiles. The comparison of both offline and online data allows
this study to analyze patterns in how these language phenomena are included by Valencian bilinguals in both conversation and in online social media posts.
CHAPTER TWO

LANGUAGE CONTACT, CODE SWITCHING AND BORROWING

2.1 Language contact and bilingualism

The assumption that languages are in contact, or rather that the *speakers* of languages are in contact with speakers of other languages (or speakers of the same language who speak it differently than they do), is a basic assumption in contact linguistics. The effects of language contact cannot be accurately predicted with one linguistic feature. A useful feature in measuring the rate of change is to analyze the rate of code switching found in a community. This measurement situates bilinguals at the center of language contact and change. “Change” in this case refers to the “gradual replacement of one or more competing variants by another” (Poplack et al. 2011). Variation is the synchronic manifestation of synchronic change, however, variation is not necessarily the same as change. The degree of bilingualism in a community is an important factor in evaluating phenomena such as CS. There are however many ways in which bilingual speakers can be bilingual as well as varying degrees of bilingualism in a community. Romaine (1989) notes that speakers may be competent in terms of grammar and phonology but lack literacy, that a bilingual can have lexical gaps or a lack in communicative competence in the second language, or that a bilingual speaker may have communicative competence but lack full control over their use of forms. Dorian (1982) notes that in Scotland some speakers have minimum skills in expressing themselves in Gaelic but have excellent receptive competence of it.

Ferguson discusses diglossia where there is a superposed ‘high’ language (H) and another variety, the ‘low’ (L) variety. He defines diglossia as:

[A] relatively stable language situation in which, in addition to the primary dialects (which may include a standard or regional standards), there is a very divergent, highly
codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation” (Ferguson 1971: 16)

In diglossic situations, there is a functional specialization of H and L. In cases where H is a written and/or formal language and L is a vernacular and the two are genetically related, this is sometimes called digraphia. Romaine (1989) elaborates on the issue of domains, originally discussed by Fishman, Cooper, and Ma (1971), as “an abstraction which refers to a sphere of activity representing a combination of specific times, settings and role relationships” (29). Common domains can be family, friendships, religion, employment, and education, and each domain may invoke the use of one language or another in a bilingual speaker’s mind. Hoffman (1971) classifies the following as possible domains: the home, the neighborhood, the school, the church, and the workplace. Maintaining the domains in which a minority language is used is important to maintaining bilingualism. Romaine (1989) notes:

There is an almost one-to-one relationship between language choice and social context, so that each variety can be seen as having a distinct place or function within the local speech repertoire. Where such compartmentalization of language use occurs, norms of code selection tend to be relatively stable. Although speakers in diglossic situations must know more than one code, only one code is usually employed at any one time (111).

Linguistic borrowing and code switching are two phenomena associated with linguistic contact and bilingualism. The following sections describe the theoretical background with regard to these two phenomena. While many scholars (Casesnoves Ferrer 2010) suggest a diglossic situation in Valencia, the situation is in reality more complex.

2.1.1 Language choice

In situations of bilingualism, bilingual speakers are faced with the decision of which
language to use in which context. In situations of diglossia the culture or community largely determines the roles for each language and in which contexts each should be used. However, in situations where diglossia is not firmly established or is less firmly established there are a number of factors, which influence language selection, which Grosjean (1982: 136) summarizes below:

**Figure 1. Factors influencing language choice (Grosjean 1982: 136)**

<table>
<thead>
<tr>
<th>Participants</th>
<th>Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language proficiency</td>
<td>Location/Setting</td>
</tr>
<tr>
<td>Language preference</td>
<td>Presence of monolinguals</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td>Degree of formality</td>
</tr>
<tr>
<td>Age</td>
<td>Degree of intimacy</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>Content of discourse</td>
</tr>
<tr>
<td>Education</td>
<td>Topic</td>
</tr>
<tr>
<td>Ethnic background</td>
<td>Type of vocabulary</td>
</tr>
<tr>
<td>History of speakers’ linguistic interaction</td>
<td>Function of interaction</td>
</tr>
<tr>
<td>Kinship relation</td>
<td>To raise status</td>
</tr>
<tr>
<td>Intimacy</td>
<td>To create social distance</td>
</tr>
<tr>
<td>Power relation</td>
<td>To exclude someone</td>
</tr>
<tr>
<td>Attitude toward languages</td>
<td>To request or command</td>
</tr>
<tr>
<td>Outside pressure</td>
<td></td>
</tr>
</tbody>
</table>

In the process of language selection, a user has three options: using language A, using language B, or code switching. All of the above factors come into play as bilinguals interact with each other and with monolinguals. According to Grosjean (1982), issues such as socioeconomic status are a principal factor in language choice while Rubin (1968) proposes the issue of intimacy between speakers as particularly influential. The content of the discourse can also be influential, as noted by Fishman (1965) who observed that some topics are better handled in one language than another because either the speaker might be competent in discussing a certain topic in only of the languages or one language might lack the necessary vocabulary for a given topic.
Hamers and Blanc (2000) in their analysis of factors influencing language choice enumerate several principles. The *language competence principle* states that the speaker selects a code “in which the sum of the individual communicative competence of the interlocutors is maximum” (253). The *ethnolinguistic affirmation principle*, which may override the language competence principle, states that, “if the gain of choosing a code well within the competence of the speaker is perceived by him to be less than the cost... to his ethnic identity, the competence principle may not be applied and a code-divergence strategy may even be chosen” (253). The *interlocutor-perceived intention principle* states that, “[I]f the speaker perceives hostile intentions on the part of the interlocutor, whether at the interpersonal or at the intergroup level, he may refuse to converge towards the interlocutor by choosing a code other than the most effective one” (253). Also, social, situational, and personal factors such as conversational roles, topic of the conversation, the social norms and/or status of the languages being used, can influence language selection (253).

2.2 Borrowing

Haugen (1950) introduced the taxonomy for classifying various types of linguistic borrowing. First, in a general sense, it is an inclusive term that refers to a variety of types of situations, but broadly relates to the adoption of linguistic elements from a speaker’s second language into their native language. Researchers later included borrowings as one of many types of *transcodic markers*, which Lüdi (1987: 2) defines as “marques, dans le discours, qui renvoient d’une manière ou d’une autre, la rencontre de deux ou plusieurs systèmes linguistiques (calques, emprunts, transferts lexicaux, alternances codiques, etc.)”. In other words, transcodic markers cover many types of language contact phenomena, including borrowing, calques, and even code
switches. A similar approach will be seen later with terms such as *translanguaging, translingual practice, polylanguing*.

Haugen (1950) describes the nature of borrowing as involving a comparison between two elements, one from the source language and the other from of the target language, which may be similar or dissimilar to each other. Two different processes are at play. One is *importation* or similarity between forms so that “a native speaker would accept it as his own” (212). The second is *substitution*, which implies the reproduction of a form inadequately with patterns from the native language. Phonetic substitution always occurs in the importation of a form from the source language to the target language. A bilingual speaker adopts a word from the second language to the first and maintains the phonetics of the source as much as possible. As more bilingual speakers adopt the word, phonetic substitution occurs with greater frequency, and, if monolinguals adopt the word, complete importation will occur.

The concept *loanword*, the broadest of borrowing types, refers to words that speakers have imported from a source language with both the meaning and phonetic shape with minimal substitution of native phonemes while the term *hybrid* is sometimes used to refer to loanwords that are partially adapted phonemically into the target language.

There are also instances where the recipient language will modify an imported form under the influence of an aspect of its own (recipient language) morphology or lexicon. *Loanblends* refer to those forms, such as in the example given by Haugen of American Portuguese *alvarozes* ‘overalls,’ where the English *o-* has been interpreted as a prefix and replaced by the Portuguese ‘prefix’ *al-*.

*Loan translations* (or calques) is the importation of a structural pattern, “the combination of the two constituents into a compound expression with a new meaning of its own not derivable
by a simple addition of the two parts” (Haugen 1950: 214). If there is no importation of structural elements and only an importation of meaning, this is referred to as a *semantic loan*.

*Loanshifts* refer to a change in meaning of a native word without any importation of forms from a source language. A related notion is that of a *semantic extension* which goes back to the work of Uriel Weinreich (1953), which he defines as, “the extension of the use of an indigenous word of the influenced language in conformity with a foreign model. If two languages have semantemes, or units of content, which are partly similar, the interference consists in the identification and adjustment of the semantemes to fuller congruence” (48). This tends to happen in genetically related languages, where the possibility of phonetic resemblance between words is more likely.

The issue of morphemic substitution is an important one to establish, which can either be inexistent, partial, or complete, according to Haugen’s schema. The following summary will show the degree to which morphemic substitution can be found with the aforementioned types of linguistic borrowing:

1. Loanwords – morphemic importation without substitution
2. Loanblends (hybrids) – morphemic importation with substitution
3. Loanshifts (loan translations and semantic loans) – morphemic substitution without importation

Where there is syntactic congruence across the languages, this will facilitate adaptation, such as the case of Norwegian and English, discussed by Haugen. Some languages are not syntactically congruent, which leads to more extensive adaptation, such as some Amerindian languages that adopt Spanish loanwords from adjectives to verbs in their grammar. The above categories show how the recipient language adapts foreign words into its lexicon.

The process of borrowing is complex. Bilinguals are often the initiators of borrowing. A newly borrowed form is usually in parallel with a monolingual form and does not necessarily
have to be adopted by all bilinguals. Prestige is an important factor in borrowing and community and speaker attitude toward the source language affects the level of diffusion of borrowed words. Sociocultural factors can present an important obstacle to borrowing. Determining the source language of a borrowed term can be difficult. Pfaff (1976) discusses how to determine the status of a borrowed word. One would have to know if the base language had an equivalent, its use in the community and if the individual knows it.

Recent studies have continued in further detailing how bilingual speakers incorporate borrowings. Lexical borrowing is the integration of words from $L_a$ in $L_b$ in which the item adopts, “[T]he (morphosyntactic) properties of the language into which they were borrowed (the recipient language), following its morphological requirements and adopting its word order, while retaining none of their donor-language identity other than their etymology, and optionally, some elements of phonology” (Poplack 2015: 919). Borrowing involves optional phonological adaptation into the recipient language.

Nonce borrowings (Poplack et al. 2012; Sankoff et al. 1990) are lone insertions that do not meet the criteria of recurrence and diffusion to be classified as established borrowings. They are content words that assume the morphosyntactic and phonological identity of the recipient language. However, nonce borrowings do not occur frequently and are not widespread. This ability of “on the spot” borrowing implies bilingual competence (Poplack 2015). The Nonce Borrowing Hypothesis (Poplack et al. 2012; Sankoff et al. 1990) postulates that bilinguals engage spontaneous borrowings via the inclusion of nonce borrowings. The criterion used to distinguish established loanwords from nonce borrowings is frequency. All borrowings are morphosyntactically integrated, while they are optionally phonologically integrated. Nonce
borrowings are infrequent in conversation, meaning that an individual form occurs infrequently, the phenomenon of nonce borrowings is a frequent occurrence in bilingual discourse.

2.3 Code switching

Code switching refers to “the mixing of two or more languages in discourse” (Poplack 2015: 918), is a linguistic behavior found in bilingual speech (Grosjean 1982; Gumperz and Hernández Chavez 1972; Kachru 1978; Pfaff 1979; Poplack 1980; Sánchez 1983; Scotton and Ury 1977). Bilinguals for a variety of reasons, consisting of contextual triggers, discursive function or in identity formation, use CS. CS can either be emblematic or spontaneous. Emblematic switching may include “fixed phrases of greeting or parting and discourse markers such as: oye, fíjate que, you know, pero (Grosjean 1982). Spontaneous CS involves “two (or more) languages… spontaneously brought to bear in a single utterance with no change in interlocutor, situation, or even topic, i.e. with no external trigger” (Poplack 2015: 918).

2.3.1 Linguistic factors in CS

Studies on CS have focused on different aspects of CS, including grammatical considerations, discursive or pragmatic functions, and socio-political aspects of how CS fits socially in a particular speech community. The inclusion of lone noun insertions was the most frequently found insertion in Pfaff’s (1979) analysis of Spanish-English bilinguals in the US. Multiword insertions occurred at a significantly less frequency.

first refers to a constraint that permits codes to be switched after any constituent in discourse, provided that the constituent is not a bound morpheme. The second constraint allows for code switches 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 structures of the two languages map onto each other.

In terms of a hierarchy of switchable constituents, Poplack (1980) found that full sentences are the most frequently switched elements, followed by switches occurring at constituent boundaries, followed by switches within major constituents such as within an NP. In terms of intrasentential switching, Poplack (1980) found that NPs are the most commonly switched elements. Poplack and Sankoff (1988) list some possible switch sites:

1) Between subject NP and VP  
2) Between verb and object NP  
3) Between auxiliary and verb  
4) Between preposition and NP  
5) Internal to the prepositional phrase  
6) Around coordinate and subordinate conjunctions

Woolford (1983) proposes that the presence or absence of an overlap in two monolingual grammars at the level of phrase-structure rules affects CS use. A speaker relies on phrase-structure rules of both languages when constructing a phrase-structure tree at the cognitive level. When the grammatical rules of the languages converge, the speaker can fill categories (NP, VP, etc.) freely from either language. When the categories are created by a rule that exists only in one language, it must be filled with items from that lexicon. The two lexicons of the languages remain separate in the mind of the speaker. Switching takes place at categories that speakers perceive to be congruent.

Poplack and Sankoff (1988) list four strategies that bilingual speakers use when employing CS in order to avoid producing ungrammatical utterances. First, speakers use ‘smooth
switching’ at equivalent sites. Second, they use constituent insertion by inserting a grammatical constituent in one language at an appropriate point for that type of constituent in a sentence of the other language. Third, speakers use flagged switching such as pauses, hesitation phenomena, repetition and metalinguistic commentary, which signals to the interlocutor(s) that a switch is forthcoming and interrupts the flow the sentence. Fourth, speakers use nonce borrowings, which are ‘on the spot’ linguistic borrowings which are adapted to the receiving language by the speaker.

Poplack (1980, 1981, 1987) describes three levels favored by bilinguals for combination of two languages. The first type is a tag, which “can typically be freely inserted anywhere in the sentence with few if any syntactic repercussions” (Poplack 2015: 918). This is a well-established pattern that functions as a way of establishing ethnic solidarity. Tags can be used by speakers who are not proficient in Spanish but pertain to the group. The second type is referred to as intersentential, consisting of a full clause or sentence constituent and syntactic requirements of either language are not a concern. Finally, intrasentential switching refers to the mixing of languages “within the confines of a single sentence or constituent” (Poplack 2015: 918). Intrasentential switching consists of two types of other-language material. First, is the inclusion of lone content words from $L_a$ in $L_b$. The other is the inclusion of multiword fragments from $L_a$ in $L_b$. Lone insertions are typically nouns but can include other content words and, less frequently, grammatical elements.

Nortier (1990: 158) proposed the distinction between the base language, i.e. the dominant language of an entire conversational episode and the matrix language, i.e. the dominant language of a sentence or similar unit. Myers-Scotton (1993, 1995, etc.) proposed an alternative model, the Matrix Language Frame (MLF) model, challenging Poplack’s (1980, etc.) theory. This model
posits a hierarchy between languages involved in a code switch situation. The matrix language plays a more dominant role than the embedded language. The MLF model also proposes a distinction between system and content morphemes within a hierarchy. System morphemes are those such as function words and inflectional affixes while content morphemes include nouns and verbs, etc. The matrix language (ML) provides the system morphemes in a phrase with the possibility of an embedded language (EL) ‘island’ – i.e. constituents only in the EL and the EL will be content morphemes. Determining the matrix language of a sentence, then, becomes a crucial task in an insertional model of code switching. There are statistical models of determining the matrix language, i.e. counting the number of words from each code, and competency-based models, i.e. assuming that the speaker will use the language he or she is most proficient in as the matrix language (ignoring cases of stable bilingualism).

Gardner-Chloros (1995) conceptualizes CS in a very different way, questioning the MLF framework, stating, “[code switching] should be considered as a much broader, blanket term for a range of interlingual phenomena within which strict alteration between two discrete systems is the exception rather than the rule” (68). She states that code switching can be found in many environments reflecting either monolingual or multilingual norms, suggesting that code switching is a reflection of the linguistic preferences of a community. She criticizes the binary implications of the term ‘code switching’,

“That at any given moment speakers are either operating in one mode or in another, which is clearly distinguishable from the first. This is an oversimplification. The type of language mixing which occurs in a given setting depends on a number of factors including the relative prestige of the varieties in question and the extent to which they are considered separate, identifiable languages” (70).

She also notes that it is rare in a contact situation for only one type of interlingual behavior to be found. She argues that the linguistic varieties involved in a code-switching
situation are non-discrete. They are only functionally separate depending on the proportion of the population that is multilingual and the degree of separateness of the domains where they are used (72). She criticizes the MLF model and questions whether speakers easily distinguish between languages in speech. Her suggestion is that “speakers can simply let down the mental barriers between the two languages at various different levels” (71).

Poplack (2000) later noted that the degree of bilingualism of the individual speaker affects the degree of code switching. In other words, a speaker who is a ‘balanced bilingual’ will have a preference for intrasentential code switching. Non-balanced bilinguals, those who have acquired the second language later in life, will prefer ‘emblematic’ switching, also called tag switches. Poplack notes the discursive function of emblematic or tagged switches in achieving interactional effects and the function of intrasentential switching as constituting a discourse mode that does not have any particular social motivation but instead forms part of the linguistic repertoire of a speech community (2000: 254-255). Lipski (2005) also noted the pattern of tag switching, found among US Latinos through the insertion of Spanish tags into English discourse, “prosodically set off from the remainder of the sentence [that] do not form part of the basic syntactic structure of the sentence” (Lipski 2005: 7).

Muysken (2000) suggested three types of code switching models (referred to as ‘code mixing’ by Muysken). This eliminated the need to find one pattern to describe all bilingual code switching patterns. Muysken proposed three modes: insertional, alternational, and congruent lexicalization. ‘Insertional’ models refer to the type proposed by Myers-Scotton, while ‘alternational’ refer to the type proposed by Poplack. Congruent lexicalization refers to the model found in varieties that share the same language but differ in vocabulary. Linguistic and extralinguistic factors can predict the code switching pattern found in a community.
2.3.2 Extra-linguistic factors in code switching

Many other researchers have considered various extralinguistic factors affecting the use of code switching in various bilingual communities.

Blom and Gumperz (1972) proposed metaphorical and transactional switching in terms of the pragmatic or discursive functions of CS. Transactional switching occurs when CS is controlled by components of the speech event such as topic or participants. Metaphorical switching concerns the communicative effect that a speaker wishes to convey to the other participants. Gumperz (1982) considered code switching as a linguistic tool or communicative option available to bilingual speakers, similar to style shifting available to monolingual speakers. He notes the informal nature of code switching and notes its presence in speakers’ discourse regardless of their positive or negative attitude towards it. Jacobson (1990) noted that within English-Spanish bilingual groups in the American Southwest, those of higher socio-economic status were more likely to use English or switch to English in bilingual discourse. However, Poplack (1980) maintained that it is the role of code switching in a particular community that gives it significance and that not each individual switch will have pragmatic significance.

Gumperz (1982) outlined some of the discourse functions of CS. First, it is used in quotations, or reported versus direct speech, where a speaker decides to report what another speaker said using the exact language of the original speaker. CS can be used to mark interjections or as sentence fillers. Another discourse function of CS is to reiterate what has already been said. Gumperz also notes the function of CS to qualify the message of a speaker. In this way, a topic is introduced in one language and comments are added in another. CS can also
be used to address the recipient of a message or to mark varying degree of involvement of a speaker in the message. The use of CS as marking ‘we’ vs. ‘they’ distinction in code choice was noted:

The tendency is for the ethnically specific, minority language to be regarded as the ‘we-code’ and become associated with in-group and informal activities, and for the majority language to serve as the ‘they-code’ associated with the more formal, stiffer and less personal out-group relations (Gumperz 1982: 66).

The contrast between codes is “meaningful and can be interpreted by participants, as indexing (contextualizing) either some aspects of the situation (discourse-related switching) or some feature of the codeswitching speaker (participant-related switching)” (Auer 1999:310).

Discourse-related switching occurs in contexts where there is a preference for one language and the identification of the language of interaction is easily determined. The switch, then, signals the “‘otherness’ of the upcoming contextual frame and thereby achieves a change of ‘footing’” (312). In this sense, CS can be a ‘group style’ and subject to normative constraints within a speech community.

Grosjean (1982) presents the following table of reasons a speaker might choose to CS in a conversational setting.

**Figure 2. Reasons for CS (Grosjean 1982)**

- Fill a linguistic need for lexical item, set phrase, discourse marker, or sentence filler
- Continue the last language used (triggering)
- Quote someone
- Specify addressee
- Qualify message: amplify or emphasize (‘topper’ in argument)
- Specify speaker involvement (personalize message)
- Mark and emphasize group identity (solidarity)
- Convey confidentiality, anger, annoyance
- Exclude someone from conversation
- Change role of speaker: raise status, add authority, show expertise
Heller (1995, 2007) discusses the practice of code switching from a politics of language point of view. In this sense, code switching “illuminates language politics” (159) when linked to other language practices. Code switching is a practice whereby users demonstrate their linguistic and cultural knowledge. However, this practice is not necessarily available to all speakers, in which case the nature of individual linguistic repertoires must be understood. As it reflects language practices, code switching is a political decision that reflects processes of power. Because of the political nature of code switching, it depends on the linguistic resources available to users to incorporate into their linguistic repertoire.

Milroy and Li (1995) analyze code switching from the social network point of view. A social network may be seen as “a boundless web of ties which reaches out through a whole society, linking people to one another, however remotely” (138). It can have weak or strong ties, depending on the level of interaction between the participants. They note the use of code switching as a preference marker which gives expression to acceptances and agreements or refusals and disagreements (150). It can be used for repair, word finding, self-editing, repetition, clarification, confirmation, etc. (151). Finally, it can be used as a pre-sequence, “a type of conversational structure which prefigures or clears the ground for a later interactional episode” (152).

Individuals interact with societal rules about language use. In her study of French and Alsatian in Strausbourg, Gardner-Chloros (1995) notes the ‘unmarked’ code switching option, which means that all will be able to understand regardless of code choice (82). She understands code switching as an individual ‘branding’ opportunity and notes the remarkable differences in frequency among speakers. Some speakers show intense grammatical mixing, in what she styles
‘unconstrained code switching mode’ (85), while others only make a symbolic gesture towards this activity by including tag switches at sentence boundaries.

Auer (1995) conducted a conversational analysis of CS and views it via the pragmatic domain (132). He studies CS from a view of contextualization, in which all of the activities go into the interpretation of an utterance (123). Auer (1995) lists the contexts in which code switching is most likely to occur:

- Reported speech.
- Change of participant constellation, particularly addressee selection.
- Parentheses or side-comments.
- Reiteration, that is, quasi-translation into the other language, for purposes of emphasis, clarification, attracting attention, etc.
- Change of activity type, also called “mode shift” or “role shift.”
- Topic shift.
- Puns, language play, shift of “key.”
- Topicalization, topic/comment structure. (120)

Contextualization cues only have meaning in context and, therefore, lack independent meaning. A conversational analysis relies on the linguistic repertoire of the community.

Sebba and Wootton (1998) discussed that this “we/they” distinction is not always useful in every bilingual situation. Additionally, the “we” and “they” must be identified in each community. McConvell (1988) views code choice as a way of expressing the social arenas in which the speaker wishes to place his relationship to the interlocutor. Giacalone-Ramat (1995) raises the issue of typologically similar languages and standard/regional CS. She notes that bilinguals have a ‘go-to’ preferred code (51), which they default to or revert to if having already switched to another code. In her study, it is largely an individual choice. However notions of group style could extend to include a communal preferred code. With regard to the direction of switching, she states that it “is not relevant in terms of social meaning or function in our data, because it depends on the preferred code: we expect that speakers using the dialect more
frequently will switch to Italian for contrasting effects, and vice versa” (52). This notes the contrastive quality of code switching and its relation to the emotional attitude of the speaker.

The role of pragmatics in code switching has been analyzed in a wide range of communities and languages such as Ewe-English in Ghana (Amuzu 2012), Spanish-Basque in Spain (Barredo 2001), Japanese-English (Nishimura 1995), Hebrew-English (Maschler 1991, 1994) and in many other studies.

2.4 Code switching vs. borrowing

The question then arises, after having considered both borrowing and CS, how are the two distinguished? The Nonce Borrowing Hypothesis (Sankoff et al. 1990) noted that the presence of lone insertions in a bilingual corpus is likely borrowings, rather than CS, either established or nonce, i.e. spontaneous borrowings from bilingual speakers. However, this does not imply that single word CS is impossible but, rather, rarely found in data (Poplack 2011: 2). According to the Nonce Borrowing Hypothesis, for a lone insertion to be considered a case of CS, it must show that it is retaining the morphosyntactic properties of the original language and not being integrated into the recipient language. Those lone insertions that behave according to the grammar of the host language are cases of single-word CS, which is theoretically possible, yet rare in data.

In opposition to the Nonce Borrowing Hypothesis (Sankoff et al. 1990), other researchers have proposed a less clear boundary between CS and borrowing (Gardner-Chloros 1995, 2004, 2009; Myers Scotton 1992). Myers-Scotton (1992) proposes a continuum so that they are not distinct phenomena, yet neither are they exactly the same phenomenon. She distinguishes between ‘cultural B forms’ or cultural borrowing and ‘core B forms.’ The first refers to ‘objects
or concepts new to the ML [matrix language] culture” (28), while core B forms are borrowed not to fill a lexical or cultural gap, but because “certain types of contact situations promote desire to identify with the EL [embedded language] culture or at least aspects of it” (29). She suggests that core B forms start out as code switching forms and “are repeated until gradually they become B forms” (30). Myers-Scotton (1993) posits that frequency of occurrence is the best criterion to distinguish single-item borrowings from code-switches. Whereas borrowed forms may appear rather frequently in a large corpus, code-switches may appear only once. Poplack and Sankoff (1984) propose that not only frequency of use but also native language displacement, morphophonemic and syntactic integration, and community acceptability should be considered in identifying single words as instances of code-switches versus borrowings. They suggest that as more of these criteria are met we can be more confident that we are dealing with loanword adoption rather than some other phenomenon.

Gysels (1992) proposed that there was no categorical distinction between borrowing and CS. Later, Gardner-Chloros (1995) proposed a similar assertion, maintaining that borrowings and CS are not binary options. She noted that both loans and code switches can be morphologically and phonologically integrated, or not; that both loans and code switches can fill ‘lexical gaps’ in the surrounding languages or add an alternative to a native equivalent. And that all grammatical categories are potentially borrowable, although nouns are most often found as both code switches and loans (73-4). In this sense, a loan is somewhat like an institutionalized code switch and can universally be styled as a lexical borrowing when monolingual speakers appropriate it to their speech.

Blas Arroyo and Tricker (2000) discuss distinguishing lexical borrowing from code switching in typologically similar languages (Spanish and Catalan) by analyzing sites of conflict.
between the competing grammars. They applied the comparative variationist method (Poplack & Meechan 1998) to the context of Valencian and Spanish. In presenting a case for maintaining a differentiation between code switching and lexical borrowing, they hypothesize that:

If the patterns of grammatical variability of nouns of Spanish origin in a Catalan context coincide with Catalan nouns but not with Spanish nouns in a Spanish context, we shall have to conclude that, leaving aside their etymology, their grammar corresponds to Catalan. That is to say, in taxonomical terms, they are lexical borrowings and not code-switches… If the variability of these lexical elements [i.e. Spanish nouns in Catalan discourse] coincides with that observed in Spanish nouns in a monolingual Spanish context or in unambiguous code-switches, we shall have evidence that the linguistic mixture in our corpus occurs basically in the form of language alternation (Blas Arroyo and Tricker 2000: 107).

They go on to compare “(1) Spanish nouns in a Spanish linguistic context… (2) Spanish nouns in a mixed linguistic context… and (3) Catalan nouns in a Catalan monolingual context” (109), using a comparative approach to establish borrowing vs. code switching for isolate nouns. They concluded, “ambiguous nouns present a pattern of grammatical variability different from that of nouns of Spanish origin, whether these appear in a monolingual Spanish or in a code-switching context. Second, the grammar of ambiguous nouns, in contrast to Spanish nouns and code-switching nouns, resembles that of Catalan nouns” (133); ambiguous nouns behave much like established borrowed material. They present grammatical integration as a key test in determining whether an ambiguous noun is a switched element or a borrowed element.

Discourse markers are a frequent site for code switches and may serve as triggers for it (Pfaff 1982, Brody 1987). Several researchers have studied the path taken by discourse markers moving from being instances of code switches to ones of borrowings. Brody (1987) suggested that Spanish particles in Tojolabal Mayan originated as code switching. She refers to the use of such markers as emblematic and claims that they constitute indigenous speakers’ “foray into the prestige language” (509). Poplack and Sankoff (1984) considered native language displacement,
morphophonemic and syntactic integration and community acceptance as major factors in
determining the status of a discourse marker as either a case of borrowing or CS. Myers-Scotton
(1993) proposed analyzing their frequency, in conjunction with her overall approach to
distinguishing borrowings from CS, as a primary way of coding discourse markers. Myers
Scotton (1993) hypothesizes that high frequency borrowed forms probably enter the language
initially as code-switches and then gradually become integrated into the language as borrowings
as they are used more and more frequently alongside the native forms they often duplicate.
Maschler (1994) studied Hebrew discourse markers in mostly English speech in order to support
a theory of metalanguage in bilingual conversation. Maschler (1994) argues that switched
discourse in general may function in a metalinguistic fashion, that it offers information about
language rather than communicating other types of information from the narrational world.
Discourse markers serve as boundary markers framing items such as different verbal activities.
They present shifts in context or new components in a narrative. Goss and Salmons (2000) have
studied English discourse markers in German, noting that English-language discourse markers
were introduced initially into German speech through emblematic code switching. English and
German dual discourse marker system existed until the German forms began to die out. Torres
(2002) notes that discourse markers can be interpreted as cases of CS or borrowing. Many
researchers have noted the frequent borrowing of common Spanish discourse markers, such as
_bueno_, in regional languages spoken in Spain, such as Catalan and Galician (Cuenca and Marín
2007; González 2004; Thomas 2008; Vila i Moreno 1998).

2.5 Conclusion
This chapter has presented a summary of previous studies analyzing language contact phenomena, specifically code switching and lexical borrowing. It situates the current study within previous studies on CS. The framework proposed by Poplack (1980, 1981, 1990, 2015) is used to analyze the data in this study, however, differing points of view were also offered in the analysis of previous studies. It showed the important work of other researchers on typologically similar languages, such as Giacaone-Ramat (1995) who notes the importance of an individual preferred code and the contrastive function of code switching between typologically similar languages. It presented the important considerations of Gardner-Chloros (1995) in considering the differences between borrowing and code switching. It also presented the important work on the extralinguistic factors that can influence language choice by noting the sociopolitical function of code switching (Heller 1997, 2007) and the importance discourse functions of code switching (Gumperz 1982, Auer 1995). Many of these considerations will be relied on in the analysis of data in this study, particularly with regard to the codification of various language phenomena.
CHAPTER THREE

POLYLANGUAGING, CMC, TWITTER

3.1 Translanguaging and polylanguaging

This study relies on the concept of polylanguaging, particularly when analyzing online data. Polylanguaging refers to “the use of features associated with different ‘languages’ even when speakers know only a few features associated with (some of) these ‘languages’” (Jørgensen et al. 2011: 33). Polylanguaging also has something to say about competency and an alternation of codes. Jørgensen et al. (2011) note these diverse language practices in social media sites. Users who polylanguage are those who include elements in their posts from a variety of languages even though they might not be speakers of those languages. Jørgensen et al. (2011) offer the following example.

Har købt the equipment, skal bare finde tid til at lave en spektakulær én kun tje dig morok, den skal være speciel med ekstra spice :P, sorry tar mig sammen denne weekend! insAllah

*Translation*: have bought the equipment, must just find the time to make a spectacular one just for you morok [Turkish = old man, father], it must be special with extra spice :P, sorry pull myself together this (weekend)! insAllah [Arabic = if God wills it]

The user is a Danish speaker who includes these elements from Turkish and Arabic in a post, although the user does not speak those languages, as reported by Jørgensen et al. (2011).

Polylanguaging derives from the concept of *language*ning, the use of language, not a language (Møller and Jørgensen 2011: 1). In this sense, polylanguaging implies that speakers can learn partial features of languages without learning the entire system itself. This leads to speakers being able to incorporate these features, such as in the example above. The conceptualization of languages as a set of linguistic features allows polylanguaging to occur so that a speaker does not need the entire system to be able to use only one feature.
Languages are constantly in contact and the ‘borders’ between them are often artificially imposed and have their roots in 19th century European nation making rather than being a reflection of a linguistic reality found in natural speech. This is not to say that named varieties do not have significance. The concept of distinct named languages, ‘English’, ‘Spanish’, ‘Catalan’, etc. are ideological constructions reflecting the emergence or product of European nation-states (Jørgensen, et al. 2011), when “the idea of autonomous languages free from agency and individual intervention meshed with the differentiation of peoples in terms of spiritual essences” (Androutsopoulos 2014: 2). The alternative to this is to focus on variable ways in which linguistic features with identifiable social and cultural associations are clustered together. Linguistic resources acquire labeling (i.e. ‘Catalan’ or ‘Castilian’) through social appropriation. These labels have meaning for the groups that they describe and can provide a rich source of identity for the group. However, they are not ontological realities, but are social constructs that serve a specific purpose (Androutsopoulos 2014: 16). Speakers store all of these resources in one linguistic repertoire. Related to the idea of one linguistic repertoire is the idea of speaker proficiency. In view of such a conceptualization, one might argue that what speakers have is an integrated proficiency which allows them to access a diverse set of linguistic resources at their disposal. Jørgensen et al. (2011) frame ‘native speaker’ in terms of rights that speakers have with respect to a set of linguistic resources (31); consequently, language learning is a process of being accepted by those with rights. Recent studies on polylanguaging have implications for the study of CS due to the conceptualization of a linguistic repertoire instead of two grammars in function. This conceptualization of language use prioritizes multilingual practice as the norm, rather than the exception. It highlights the irregularity of monolingualism in the sense that CS and borrowing are not irregular phenomena in the use of language.
The notion of superdiversity also tears down boundaries between linguistic communities vis-à-vis contact zones as it notes the polylingual nature of some forms of CMC. Originally superdiversity refered to offline realities in urban centers, particularly in Britain (Androutsopoulos and Juffermans 2014) where there has been an increase in migrants “not only in terms of nationality, ethnicity, language, and religion, but also in terms of motives, patterns and itineraries of migration, processes of insertion into the labor and housing markets of host societies” (Blommaert and Rampton 2011: 1). Superdiversity emphasizes the importance of new technologies, which encourage global flows of people and ideas. In this presumed new reality the boundaries between languages are not easily discernible. Contact zones are the “spaces where diverse social groups interact” (Canagajarah 2011: 26), where there occurs “social interaction between groups positioned differently in power” (Pratt 1991: 30). Läppenen and Peuronen (2012) argue that “the translocal Internet has also become a linguistic contact zone in which multilingual resources and repertoires can turn out to be crucial capital for successful communication, action, and interaction” (389).

What these theoretical reconsiderations have in common is that they privilege practice over function, where “codes attain grammatical status and also begin to signify certain meanings by convention” (28). However, the translingual framework does not deny the social power of grammatical forms via the construction of linguistic norms, but rather offers a focus away from a discussion of speaker competency in favor of a consideration of the linguistic repertoire of a speaker, ignoring issues of origin, upbringing, proficiency and language type. They reference an individual speaker’s variable grasp of a plurality of styles, registers, and genres, where the notion of competency is less important than a speaker’s identification with a set of linguistic features. They suggest the terms sensibility or structure of feeling as alternatives to competency. This
orientation away from competency is particularly important in relation to polylanguaging, which specifically proposes that speaker competency is not important with regard to their access to using many linguistic features, and that speakers will take on linguistic forms and/or discard them as they affiliate and disaffiliate with various groups throughout their lives. In this sense, communication itself also becomes more fluid as linguistic convention or structures are a resource available to speakers and the notion of meaning is a process of “here-and-now projection and inferencing” (4). The notion of indexicality is equally important, which denotes the connotational significance of signs and notes that registers or codes carry associations that are relevant to specific activities.

Many have criticized the suggestions made by researchers claiming superdiversity as a factor in multilingualism. Flores and Lewis (2016) suggest an alternative framework for viewing emerging linguistic practices, such as polylanguaging. This alternative point of view rejects the tendency to classify all linguistic phenomena, which is reminiscent of earlier work in CS (Gardner-Chloros 1995) and others who questioned the role of proficiency with language use (Auer 2000). Flores and Lewis (2016) also highlight the issue of recognizing how speakers themselves think of language phenomena. For instance, when a speaker inserts an item from another language does the speaker themselves consider that an act of inserting an element from another language? These are important considerations that should guide the discussion on classifying language practices, taking into consideration situational and contextual settings.

3.2 Social media

As noted by boyd [sic] and Ellison (2008), social network sites are “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system,
(2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” (210). They avoid the term ‘networking’ due to its connotation as a relationship initiation activity, which is not the purpose of social network sites. According to them participants create a visual representation of their social networks in social network sites rather than initiating new relationships. boyd (2011) also describes the notion of networked publics to describe social network sites because they are “(1) the space constructed through networked technologies and (2) the imagined collective that emerges is a result of the intersection of people, technology, and practice” (39).

Important components of social network sites include the ability to see and explore the friends’ network as well as the ability to leave messages on friends’ profiles. The authors note that typically there is some offline connection between social media friends, even if it is a weak personal connection.

boyd (2011) lists some of the basic features of all social network sites. First, there is the profile, the space where users “must determine how they want to present themselves to those who may view their self-representation or those who they wish might” (43). The profile also allows for conversation and the sharing of ideas. While the profile is potentially ‘truly public’, that is available for all to see, it is commonly limited in some way via the privacy settings allowed by the social networking site, thus making it ‘semi-public’ (43). Another aspect of social network sites is the ‘friends list’ (called followers on Twitter), in which participants “articulate who they wish to connect with, and confirm ties to those who wish to connect with them” (43). Usually, the friend status must be confirmed before a user’s content will appear on the other’s profile. The friend’s list is not only the closest friends but include a wide spectrum of contacts, which boyd (2011) styles as an “articulation of a public” (44).
Androutsopoulos (2014) defines a social network as a set of semiotically materialized, interactive connections among human participants (62). In contrast, a social networking site, or social network site (boyd and Ellison 2008: 210) is a “bounded communication system that enables the formation of social networks among registered participants and affords them various tools for representation and interaction” (62). Androutsopoulos (2014) refers to the user as ‘ego’ and notes that the profile page is where ego’s activities are displayed as well as where the ‘friends’ respond to them (it is important to distinguish between social media followers, sometimes called ‘friends’ and natural friends because the former does not necessarily imply a personal relationship). By contrast, the newsfeed (with various names and avatars in different social media sites) is a display of all the activities of ego’s ‘friends’. These varying relationships in social networking sites “tend to give online expression to existing offline communities” (Seargeant et al. 2012: 514). Androutsoupolous (2014) also distinguishes two types of audiences, first, the networked audience is ‘friends’ who share a connection with ego and not to ego’s other ‘friends’ necessarily but can share independent relationships with each other. This contrasts to ego’s imagined audience, or who ego perceives to be reading his or her posts.

While various social network sites share many features, there are some differences between them. Twitter use is defined as a form of microblogging service, that is, “a form of blogging that lets you write brief text updates (usually less than 200 characters) about your life on the go and send them to friends and interested observers via text messaging, instant messaging (IM), email or the web” (Wikipedia). Java et al. (2007) note the communicative need fulfilled by microblogging,

“By encouraging shorter posts, it lowers users’ requirement of time and thought investment for content generation. This is also one of its main differentiating factors from blogging in general. The second important different is the frequency of update. On
average, a prolific blogger may update her blog once every few days; on the other hand a microblogger may post several updates in a single day” (57).

They note the following characteristics concerning Twitter use. The Twitter community is composed of people who have common interests and share with each other about personal feelings and life experiences (61). There exist the following general themes with Twitter posts (tweets): (1) daily chatter, (2) conversations, (3) sharing information, and (4) reporting news (62). User roles include: (1) information source, (2) friends, and (3) information seeker. (1) is the type of user who has a large following but this does not necessarily mean they are hyperactive. They post news and/or other valuable information that attracts (3) who is mostly inactive in producing their own contributions. (2) implies some personal connection between Twitter followers.

Twitter as a type of microblogging is a social media outlet that prompts users to answer the question, “What are you doing?” thus “creating a constantly-updated timeline, or stream, of short messages that range from humor and musings on life to links and breaking news” (Marwick and boyd 2010: 116). There is no necessity for reciprocity among users and often there is no social expectation of it. Lee (2011) conducted a study of micro-blogging and status updates on Facebook. She defines micro-blogging as the “writing of short messages on the web designed for self-reporting about what one is doing, thinking, or feeling at any moment” (111). Consequently, users post both microblogs and tweets for various reasons, such as to share information, interact with others, and to discuss current events. This type of communication should be viewed as a creative cultural process. Lee uses the following taxonomy to distinguish types of status updates found in Facebook:

i. What are you doing now?
ii. Everyday life
iii. Opinion and judgment
An important point made by Lee is that status updates are delayed communication, meaning that users do not feel a need to mix languages as they do in conversation due to the non-instantaneous feedback that they will receive from other users.

3.3 Multilingualism on social media sites

Research on CS on the Internet from a discourse analysis perspective has shown its usefulness in *identity and social practice*, which is created and negotiated through the use of semiotic resources. Style shifting and code switching can be used by authors/speakers to negotiate interpersonal relations. Hinrichs (2006: 86) studied the use of Jamaican Creole and its use in CMC. He found that switches to Creole added a ‘communicative value’ in constructing identity. They note the importance of language in identity construction in CMC due to the lack of other ‘visible’ markers such as race, gender, or social class (394). Dorleijn and Nortier (2009) note the hybridity found in the Internet between speaking and writing, which is sometimes classified as a ‘third medium’ of communication (Crystal 2006).

The whole business of ‘multilingualism’ on the Internet can be a tricky one, as Leppänen and Peuronen (2012) state, “multilingualism on the Internet can refer to the practices of multilingual Internet users and the ways in which they draw on and use resources provided by more languages than one in their CMC” (385). They address the issue of language choice as a *semiotic strategy* that can be used by participants to “distinguish themselves and to create voices
of their own in a multilingual online setting” (390). These language choices can shape individual identity and the notion of community.

Eleta and Golbeck (2014) look at multilingual use and Twitter. “We see a different Internet depending on the language we use” (424). They see a continuum of increasing connections between language groups analyzed in their study or, alternatively, a continuum of increasing penetration of one language group within the structure of another. Smaller or less cohesive language groups lie at the periphery of the social graph. The proportion of English users in a network constitutes a key-influencing factor in the frequency of English use by multilingual users. The proportion of English speakers also correlates to the use of L2 by other multilingual users. Their study shows that multilingual Twitter users perceive the language composition of their network and interact with their followers accordingly. Alternatively, they suggest that the language choices of multilingual users might attract an audience of like-minded users. However, the relationship between these two tendencies is likely a self-feeding cycle (431). However, with regard to the use of English online, Lee and Burton (2012) found that English is a lingua franca online although overall content in English dropped from 80 % in 1998 to 56% in 2002 (129).

Hale (2014) notes the importance of multilingualism in the Twitter network, noting that over ten percent of users engage in content from multiple languages and that these multilingual users are more active than monolingual users. He notes that English is the most widely used language on Twitter although it only accounts for less than half of the messages exchanged on Twitter. The notion that individuals will group together based on similar characteristics is central. The most active users will form the core of a network and serve as a bridging role in the community of users.
Androutsopoulos (2013) investigates the role of multilingual language practices on Facebook. He notes the importance of recent reconceptualizations of language practice by ‘bilingual’ speakers of multiple ‘languages’ (see sections above). The various terms, polylingualism, metrolingualism, and translanguaging, are all terms that reflect a shift of focus in linguistic research from linguistic systems to multilingual speakers, i.e. the speaker is the focus. Androutsopoulos (2013) adds to the steady flow of criticism of a static concept of ‘languages’ that speakers acquire and continues the re-focusing of language boundaries as fluid and flexible via the conceptualization of a ‘language’ as a set of linguistic features or resources. Accordingly, code switching is less thought of as “how distinct codes are switched and mixed”, but, rather, “how language users manipulate the resources they have available to them” (2). He also notes the usefulness of terms such as metrolingualism, polylingualism, and translanguaging for understanding the ways in which multilingual speakers include linguistic resources in their repertoire. Metrolingualism questions the correspondence between spatial territory and language. Polylingualism’s orientation destroys the link between competency and language use, in which case, users rely on language crossing, the use of elements of other languages to negotiate ethnic and class relations according to Androutsopoulos (2013).

Androutsopoulos (2013) then introduces the term networked multilingualism, which he uses as a cover term for all multilingual practices that are shaped by two interrelated processes: (1) being networked – digitally connected to other individuals, and; (2) being in the network – embedded in the global digital mediascape of the web. Networked multilingualism encompasses all practices of users with the range of linguistic resources available to them. However, networked multilingualism is constrained by the following items: (1) digital literacy repertoires (mediation of written language by keyboard and screen technologies); (2) Network resources,
and; (3) Networked audiences. (1) refers to limitations in a user’s alphabetization and social conventions for a written usage of the language, as well as orthographies and scripts, and their availability in the web. (2) refers to the access of users in the network to semiotic resources available on the web, such as Google translate, which increases the linguistic resources available for use. (3) refers to the make-up of the audience itself. The notion of context collapse comes into play, as the audience in social media is comprised of individuals who, in other contexts, are usually not addressed concurrently (i.e. friends, family, colleagues, etc.). In his study, Androutsopoulos found that students’ networked multilingualism is individualized, shaped by genre, and based on a wide and stratified repertoire. Languages are associated with particular individuals, genres, or thematic occasions. English is used in genres that are detached from the immediate circumstances.

Theodoropoulou (2015) studied the use of Greek birthday wishes on Facebook analyzing the construction of politeness on Facebook. Her data spans three years of data collection and contains an extensive corpus of birthday wishes. Her data shows that users on Facebook pave the way for linguistic politeness by thanking wishers. This is accomplished by being personally targeted, the use of Greek forms expressing ‘thank you’ and ‘may you be well’ and by attaching a first name or nickname.

Kim et.al. (2014) study the multilingual use of languages on Twitter in Switzerland and Canada, focusing on overall language use, the effect of different languages as a barrier or bridge among friends, the hierarchy of languages in multilingual societies and the language selection of multilingual users. Their study finds that monolingual users cluster together with lower levels of linguistic diversity found in their tweets and that users who tweet in the local, regional language
have more followers and friends, even with the presence of more English monolinguals in the network.

Jurgens et al. (2014) conducted a study on the presence of code switching in hashtags. Users are perhaps motivated to include hashtags in another language to increase their potential audience size or to appear as a member of the multilingual virtual community. The authors note the widespread use of hashtags on Twitter (and thus a rich source of data). They note that there are a number of phenomena often falsely classified as code switching by researchers. They suggest two criteria for establishing that an instance of multiple language use is a code switch. First, there must be sufficient text to determine the base language of the message. Second, the message must be an act of communication by the user and not a spam-like message or advertisement (56). Jurgens et al (2014) also discuss a proposed taxonomy of hashtags based on their purpose in the message. According to this taxonomy, some types of hashtags lend to multiple language use that should not be categorized as CS. For example, the use of the names of entities, applications, and other advertisement-like tags would not constitute CS. Some examples of these types of multi-language hashtags that would not be considered CS are tags such as (1) #TeenChoiceAwards, (2) #followback, (3) #iPhone, (4) #iHeartAwards, and (5) #forsale. (1) and (4) refer to an event. (2) refers to actions or functions on Twitter. (3) refers to a brand. (5) refers to an advertisement.

Jurgens et al. (2014) note that an increased rate of bilingualism decreases the frequency of including hashtags from another language but increases the overall rate of code switching when hashtags are present. The authors note the dual purpose of hashtags as a way of bookmarking particular content and also as a method of community formation through the
discussion around the topic. Users may adopt a common hashtag in another language for reasons other than code switching.

Vallejo (2011) conducted a study about the use of code switching by Twitter users from Tijuana, Mexico. She found that code switching was used by the informants for two primary functions: (1) as a way of presenting a positive self-image to followers, and (2) linguistic creativity. (1) allows the user to create an atmosphere of intimacy between follower and author and to build confidence between the two. (2) is manifested through transference and code switches and allows the author to interact in an additional way with the audience and construct the communicative situation.

Poell and Darmoni (2012) studied the effect of multilingualism in light of the Tunisian revolutions. The authors note the presence of Arabic, English, and French in the tweets of the users. They use each language to address a different public. The study focused on the idea of ‘Twitter revolution’ and showed that it was not as influential as it was reported to be in the media; however, it was important in the overall communication process.

Siebenhaar (2006) conducted a study on the use of Standard German and Swiss dialects in IRC chat rooms. He found that there is a generational difference with regard to language preference. Older users prefer standard German, while younger users are more likely to use dialectal forms online.

### 3.4 Audience Design on Twitter and language use

Much of the work more broadly on stylistic variation is particularly applicable to analyzing language use in Twitter postings given that both the audience composition is less established than in normal conversation and the speech acts are arguably more driven by issues
of the performance or projection of an identity. Bell’s (1984) *Audience Design Model*, applied initially in monolingual contexts, is perhaps one of the most useful approaches to understanding language use on Twitter. The basic premise of the audience design model is that style shifts result from a response to audience members or a change in audience membership (Bell 2001:141). Speakers design their speech around the audience members and this purposeful design of linguistic features affects all aspects of a language system, such as pronouns, address forms, politeness, as well as switching between codes (145). There is an overlapping schema of various participants in the conversation and their relationship (or lack thereof) to the speaker. Therefore, the various members of an *audience* must be clarified, given its overall importance for the speaker’s language use. Bell (1984) divides the audience into the following categories of participants: (1) addressee; (2) auditor; (3) overhearer, and; (4) eavesdropper (159). The classification of a person in a category depends on whether or the person is ‘known’, ‘ratified’, or ‘addressed’ by the speaker. The following table summarizes the characteristics of each group.

**Table 1. Audience roles**

<table>
<thead>
<tr>
<th>Category</th>
<th>Known</th>
<th>Ratified</th>
<th>Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addressee</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Auditor</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overhearer</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eavesdropper</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

What do speakers respond to in their audience? They respond to personal characteristics of the audience members, to the general style of their addressees’ speech, and the addressees’ roles for specific linguistic variables (167). In each case, the speaker assesses the preceding characteristics of the addressee and designs their style to fit with the features of the addressee.
Androultosopoulos (2014) analyzes the nature of audience design on Facebook. He modifies Bell’s (1984) audience design framework to fit the peculiarities of social media, making the addressee a member of a networked audience directly addressed by ego. Bystanders are members of the networked audience who are actively engaged in a wall exchange (on Facebook). The entire social network is an overhearing audience. Given that the act of following ego would constitute ratification in the exchange of information, Androultosopoulos does not address the status of the eavesdroppers, who would appear to be the whole social network who on Twitter have access to tweets unprotected by the security measures selected by the user.

In his study of Facebook, Androultosopoulos (2014) notes the following about modifying audience design framework for social media. Overall, the goal of each contribution is to ‘maximize the audience’. Users employ three techniques to accomplish this: (1) choice of common-denominator language; (2) replication of given propositional content in two or more languages, and; (3) refraining from linguistic resources altogether. (1) can be a language such as English, which bridges the gap in comprehension by audience members and (2) is a semiotic effort by ego to address the audience in their language in order to maximize the audience. Users can tailor contributions to specific segments of the audience, similar to what Johnson (2013) found with Welsh-English bilinguals. Androultosoupoulos (2014) notes the importance of language style and language choice as key resources in a situation of context collapse. However, it should be noted that language use patterns found on social media do not necessarily reflect offline conversational usage, although the opposite can be true, as Seargeant et al. (2012) state, “[social networking sites] tend to give online expression to existing offline communities” (514).

The Style Axiom (Bell 1984: 151), which states that “Speakers must know the linguistic resources available in their community in order to strategize as to how to use them most
efficiently. This equally applies in bilingual situations, where the linguistic resources of multiple
linguistic sets (‘languages’) are drawn upon to create stylistic variation.

Some style shift is not accounted for by the audience members but by the topic and
setting (Bell 1984: 178). Blom and Gumperz (1972: 429) note that a change in topic can trigger a
change from local to standard dialect. Setting is harder to account for in research but could affect
style. Bell (1984) challenges the notion of topic as a cause of style shift. He cites a number of
studies that contradict Blom and Gumperz and notes that style shift caused by a change in topic
might be more prevalent in unilingual (‘monolingual’) discourse, whereas bilinguals have to
adapt to the interlocutor (180). He also notes that the peripheral audience members affect style
shift in bilingual discourse far more than in unilingual discourse. Speakers associate classes of
topics or settings with classes of persons (Bell 1984: 181).

Schilling-Estes (2002) proposes the Speaker Design Approach, which criticizes the
reactionary tendency in Bell’s approach. She argues that various contexts and topics can
motivate a style shift which can be conscious or unconscious and can be short-lived. This
approach to style shift takes into consideration the following ideas.

Language and society are viewed as co-constitutive: the linguistic features and patterns
speakers use are not mere reflections of static identity, as defined by one's positions in an
existing social order… but rather are resources speakers use to shape and re-shape social
structures such as class, and gender groups, as well as their positioning with respect to
these structures and with respect to one another (Schilling-Estes 2002: 389).

According to this model, speakers do not style shift in reaction to audience members but
are active creators of their identity through style shifting. Speakers can use a variety of linguistic
resources to achieve this means. The notion of speaker agency seeks to address why speakers
make stylistic choices rather than simply which choices correlate with which situation. Schilling-
Estes (2002) also mentions the importance of style shift as a mechanism to express group
membership. Through the use of features associated with a group, a speaker expresses that they wish to be associated with that group. Additionally, sometimes a speaker may highlight particular attributes of a group instead of overall group membership as a desirable quality with which to identify through their use of style shifting.

Context collapse refers to the phenomenon where “diverse social contexts [collapse] into one” (Marwick and boyd 2011: 10). In this scenario, various groups of people such as friends, family members, and colleagues are ‘collapsed’ into the same audience with which a user must adapt in his or her audience design. It occurs when members of a social networking site “reflect different social contexts and have different expectations as to what is appropriate” (boyd 2011:30). Context collapse is a basic feature of social networking sites because they are ego-centric with all of ego’s social connections revolving around it with little regard to audience distinction. Context collapse affects the way in which the speaker accommodates the audience and as such can be examined from Bell’s (1984) audience design framework. Social networking sites create a novel type of audience due to context collapse. This audience differs from both conversational audiences and large-scale media broadcasts. As context collapse creates a situation of overlapping linguistic repertoires (due to the linguistic features associated with each user), the users must select content for each post depending on their perception of its relevance for the social context of the audience. In other words, participants construct “context interactively as they design audience anew by means of language style in each communicative exchange” (64). One important point that Androutsopoulos (2014) makes about the linguistic implications of context collapse is the notion that context collapse leads to the destabilization of language style and leads to metalinguistic negotiation.
Marwick and boyd (2010) explore various ways in which the issues of context collapse and the imagined audience affect Twitter users. The notion of audience is important to all aspects of communication but especially so with CMC. Users conceive of an audience when they post on social media, and, although the audience in social media has few limits (especially in instances of a ‘public’ profile), users typically conceive of their audience as bound in some sense (even if they do not control their privacy settings). “In the absence of certain knowledge about audience, participants take cues from the social media environments to imagine the community” (boyd 2007: 131).

How do users perceive their audience on Twitter? First, Twitter “affords dynamic, interactive identity presentation to unknown audiences” (116), which leads to constant self-presentation via tweets and conversations with other users. Often a tweet’s actual audience differs quite drastically from its imagined audience. Markwick and boyd (2010) find that users conceive of Twitter as ‘personal space’, which relates to a lack of expectation of reciprocity among users. The goal of the user is to promote a self-brand among friends and others who are not easily conceived of as an ‘audience’ by users. Users have quite a bit of control over how they can modify tweets to suit various audiences, even if they do so subconsciously. They perceive the ideal reader as someone who is similar to the author and will share their perspective. The notion of self-presentation ties into the theory of identity as proposed by Goffman (1959), who notes that identity is a continual performance. Self-presentation is a collaborative effort and leads individuals to emphasize or de-emphasize certain things in response to audience feedback, in a “dynamic, recursive process.”

Nguyen, Trieschnigg and Cornips (2015) conducted a study on the role of audience and the use of minority languages on Twitter. They found that English is used primarily to maximize
the audience, given its status as a universal language (666). Language choice additionally is affected by constraints such as audience, topic, and perspective and may have the pragmatic function of indicating a humorous or serious comment (666). They note the importance of context collapse and introduce a classification system for understanding various types of tweets:

1. Tweets that respond to another tweet
2. Independent tweets
3. Tweets with hashtags

When a user specifically mentions another user, they are more likely to use the minority language in this study, although users do not always adapt to the addressee. Hashtags in the minority language trigger the rest of the tweet to be in that language.

Johnson (2013) conducted a study of the use of Twitter by Welsh-English bilinguals. In his study, he found that English and Welsh were used roughly the same amount of times by Twitter users, although there was a wide variety in use per user. He classified tweets based on the type of contribution according to the following taxonomy:

1. Original contribution to Twitter
2. Re-tweets
3. Response to another user
4. Generic re-tweet
5. Multiple tweets of the same message (often in different languages)

In his qualitative analysis, Johnson classified tweets in the following manner, according to the language use found in each tweet:

1. Tweets with no independent text
2. Tweets in an indeterminate language
3. Bilingual Welsh and English tweets
4. English language tweets with some Welsh
5. Tweets with English but where Welsh speaking is an advantage
6. Other languages used

He found that bilingual tweets (with both English and Welsh) were used primarily to give information quickly in one tweet to accommodate both Welsh and non-Welsh speaking
audiences. The use of Welsh also has the additional symbolic effect of expressing Welsh linguistic identity rather than solely being used as an effective form of communication. Johnson also proposes that Welsh-English bilinguals might use code switching in order to avoid problems in communicating their message within the 140 character Twitter constraint. This parallels other studies on code switching that propose economy as a criterion in determining a code (Shin 2010). Johnson also notes what he terms the ‘Welsh Advantage’ which is a technique used by bilingual users to address specific content to the Welsh-speaking audience and to purposefully exclude non-Welsh-speaking audience members. One would assume that the idea of a ‘regional language advantage’ can be derived from Johnson’s ‘Welsh Advantage’ whereby speakers of minority or regional languages use the non-mainstream code to express certain information. This gives the Welsh speakers’ privileged access to specific material. In this sense, Welsh tweets are addressed to a more specific audience and English-only tweets are addressed to the larger, more general audience.

3.5 Further considerations for CMC research

Leppänen and Peuronen (2012) note the attempt of researchers to measure “how visible and accessible particular languages are” (386). They also note less quantitative methods and purposes of research on Internet language such as focusing on “the specific multilingual practices of Internet users, the motivations behind their language choices and the functions and meanings these have for them in the specific Internet contexts in which they operate” (389).

Blommaert and Rampton (2011) note some considerations that should guide research in CMC. Some of their suggestions reflect attitudes offered by other researchers, such as those
offered by Flores and Lewis (2016) in their application of an emergentist perspective in the study of code switching:

“The contexts for communication should be investigated rather than assumed. Meaning takes shape within specific places, activities, social relations, interactional histories, textual trajectories, institutional regimes and cultural ideologies, produced and construed by embodied agents with expectations and repertoires that have to be grasped ethnographically” (5).

Many researchers note the importance of analyzing CMC data applying an ethnographic approach (Läppenen and Peuronen 2012: 395), also called a is course-centered online ethnography (Androutsopoulos 2008; Kytölä 2012), which is “[the] systematic observation of unfolding communicative activities and semiotic production [that is] guided by the accumulated knowledge accrued during the observation, versatile forms of contact with key social actors” (Kytölä 2012: 121). Online ethnography can involve only an investigation of the online activities of users with little or no contact with the users themselves. Or, it can be accompanied by fieldwork in offline (or online) settings with the participants. The context of communication is as important as informant output in determining its meaning and purpose. The authors offer the following on the issue of data analysis “Analysis of the internal organization of semiotic data is essential to understand its significance and position in the world. Meaning is far more than just the expression of ideas and biography, identifications, stance and nuance are extensively signaled in the linguistic and textual fine grain (10).”

Androutsopoulos (2013) adds important observations about the challenges in conducting online data research, particularly with the lack of socio-demographic information on language users and oral data. There are some ways to alleviate the problem. Androutsopoulos (2013) suggests “researchers can take data offered by users themselves as a basis for speaker categorization” (245). Others have proposed using the number of followers as a proxy for
establishing social prestige (Kim et. al 2014; Kwak et. al 2010). Androutsopoulos (2013) notes other challenges related to CMC research. First, CMC involves primarily written text and there is generally less sociolinguistic research on writing, noted also by (Sebba 2012). Additionally, in CMC, multimodal resources are often used as frequently as text (media, pictures, emoticons, etc.) as they represent instances of pictoral non-verbal/non-written communication.

3.6 Conclusion

This chapter has analyzed research dealing with the nature of communication online. It presented key concepts related to the analysis of language online, presenting the work of key researchers in this area of CMC (Androtsopoulos 2008, 2011, 2013, etc.). Important to this study is the analysis of the language phenomenon of polylanguaging (Jørgensen et al. 2011), used as a tool to code and classifying diverse language phenomena online, particularly with regard to the lack of proficiency of an informant in a language when an informant uses elements from that language in their online social media posts. This chapter presented research discussing the nature of online social media sites and how they relate to offline realities vis-à-vis social relationships and other considerations dealing with audience membership (boyd and Ellison 2008). It presented recent studies conducted on code switching in both Twitter and in other forms of CMC (Jurgens et al. 2014, Johnson 2013, Vallejo 2010). Finally, this chapter presented unique considerations that must be taken into account when analyzing data collected from online sources.
CHAPTER FOUR
METHODS

4.1 Introduction

This chapter describes the three published oral interview data sources used in this study as well as the method applied in constructing the Twitter corpus and will outline the approach taken in analyzing the data collected. The data analyzed comes from two types of data sources that are intended to represent and capture differences between oral and non-oral communication. The sources for data regarding the former are published corpora containing written transcripts of sociolinguistic interviews and other recorded conversations. For the latter, a corpus of data was collected from Twitter. The following sections will further describe each of the data sources. First, I will discuss the components of each corpus, i.e. the number of informants, number of words, the duration of each interview and other information pertaining to each data source. Then, I will describe the approach applied in the data analysis, such as how the data was coded and other relevant issues. In the final section I will discuss issues related to the Twitter data and how it was analyzed in comparison to the quantification and analysis of the data collected from the conversational data sources.

4.2 Published conversational corpora

The corpora used for comparison in this study are published transcripts of sociolinguistic interviews constituting bilingual corpora. The data from the corpora represent a sample of Valencian Spanish in two of the geographical centers of the community. An analysis of multiple sources of data allows for a more thorough comparison to be made regarding the contact
phenomena analyzed and the inclusion of data from two geographic locations within the
Comunitat Valenciana presents a more complete picture of the language use in Valencia.
Additionally, the methods used in the data collection are different in each of the conversational
corpora. This will become more apparent in the Chapter Five discussion of the results but the
inclusion of material from sources presenting different data collection methods adds an
additional layer of complexity to the analysis and presents the potential for a more substantive
set of conclusions.

4.2.1 Corpus sociolingüístico de Castellón de la Plana (CSCDLP)

The first corpus used in this study is the *Corpus sociolingüístico de Castellón de la Plana
y su área metropolitana* (CSCDLP) (Blas Arroyo et al. 2009). Castellón is one of the largest
cities of the Comunitat Valenciana, together with Valencia and Alicante. The CSCDLP data
offers a good representation of Valencian Spanish. The CSCDLP contains 72 sociolinguistic
interviews of around 45 minutes in duration, and 575,000 words. The informants are grouped by
the criteria of gender, age, level of education, native language, and most-used language. The
transcripts provided by CSCDLP are neither phonetic nor tagged for the linguistic features
analyzed in this study. Occasionally phonetic realizations are signaled by orthographic means
such as the use of parenthesis to indicate d deletion, “para(d)o” or the use of h to mark s
aspiration.

The following section describes the informants who participated in CSCDLP. The
breakdown of informants by gender and age is found below. There is an equal representation of
informants for each of the categories.

**Sexo**

| Hombres: 36 |
Mujeres: 36

Edad
Grupo 1 (joven): 24
Grupo 2 (adulto): 24
Grupo 3 (mayor): 24

Blas Arroyo describes the manner in which individuals were classified and the requirements associated with each category. Group 1 constitutes those under the age of the 30, Group 2 consists of those between 35-50, and Group 3 consists of those older than 55. Education grouping refers to the level of schooling by measuring the number of years completed.

Finally, informants are grouped according to their language use into two categories: lengua materna and lengua habitual. Both monolinguals, passive bilinguals and active bilinguals are informants in the CSCDLP. This is an important aspect of the CSCDLP because it shows that some informants are proficient in Valencian, as cases of the use of Valencian are found primarily in the informants that are either of the categories “lengua materna valenciano” or “lengua habitual valenciano” as these are the functionally bilingual groups. Instances of the use of Valencian in the Spanish-dominant groups are higher frequency Valencian borrowings. The following table shows the breakdown of informants for this category.

- Lengua materna castellano: 33
- Lengua materna valenciano: 36
- Lengua materna castellano y valenciano: 3
- Lengua habitual castellano: 30
- Lengua habitual valenciano: 31
- Lengua habitual castellano y valenciano: 11

4.2.2 PRESEEA (Proyecto para el estudio sociolingüístico del español de España y de América)

The second corpus to be analyzed is from the PRESEEA project in Valencia, El español hablado de Valencia. Materiales para su estudio (Gómez Molina 2001, 2005). The PRESEEA
data contains 40 hours of recorded sociolinguistic interviews conducted with informants grouped according to sex, age, and socioeconomic status. The transcripts provided by PRESEEA are not phonetically transcribed. The PRESEEA corpus provides minimal tagging of Valencian material found in the corpus through the use of footnotes on the bottom of each page. For example, on page 212, an informant uses the Valencian noun *los varietats*, and at the bottom of the page a note explains “Expresión de valenciano, *los varietats*, ‘las variedades.’” However, these notes do not tag the insertions found as either borrowings or code switching and only the first instance of each form is annotated.

The PRESEEA corpus contains 72 sociolinguistic interviews for analysis, constituting a bilingual corpus. The informants come from the capital city of Valencia. The inclusion of data from both the city of Valencia and Castellón de la Plana (CSCDLP) provides an accurate portrayal of the use of Valencian in the region due to the size of these communities. These cities are also in the part of the Comunitat Valenciana where Valencian is most widely spoken. Sociolinguistic categories used in the corpus (gender, age, language status, and level of education) were available.

The PRESEEA team gives the system used to classify informant level of education. Informants are categorized using a scheme similar to that used in CSCDLP, with three groups differentiated by number of years of schooling completed.

The PRESEEA corpus contains both bilingual and monolingual informants, as is the case with the CSCDLP. There are a slightly larger number of bilinguals (52%) in the corpus. Also, when PRESEEA classifies an informant as “monolingual” this implies a passive competency in Valencian, according to the organizers. Both PRESEEA and CSCDLP include a roughly equal number of Valencian-Spanish bilinguals and Spanish dominant informants. Instances of CS
analyzed in this study are generally found in the bilingual informant group with, however, some instances of higher frequency borrowings in the speech of monolinguals.

4.2.3 Val.Es.Co

The third oral corpus used in this study is the Val.Es.Co. corpus (Valencia, Español, Coloquial). It is different in both presentation and style from the previously mentioned corpora. It began at the University of Valencia in 1990 and has the main objective of studying “colloquial Spanish.” Val.Es.Co is a compilation of written transcripts of 43 conversations from the capital city of Valencia. Given the research goal of the Val.Es.Co team is to understand colloquial Spanish, the methodology used to collect these interviews diverges from the other conversational corpora analyzed here. The Val.Es.Co corpus is comprised of secretly recorded conversations. In this sense, the Val.Es.Co corpus provides truly spontaneous data collected in a naturalistic setting.

Val.Es.Co is available on-line from the Universitat de Valencia. The interviews are not transcribed phonetically and do not contain any tagging of lone insertions as borrowings or instances of CS. However, the type of information on each informant found in the other conversational corpora is lacking from Val.Es.Co. In this case, the bilingual proficiency of each informant is not made explicit by the study.

4.3 Twitter corpus

In addition to collecting data from the aforementioned published oral interview corpora, a separate corpus was compiled from Twitter posts taken from Valencian profiles. Sixty profiles
were selected as a source for data analysis. I employed a ‘screen-based’ site of data collection without an offline or online engagement with the informants. All tweets in a given time frame were copy-pasted from the informants’ profiles to a Word document. The ‘friend of a friend’ method was used to find Valencian Twitter users. I consulted the profile of a personal acquaintance (but did not collect data from this user) and, from this profile, I searched for other Valencian users, using the acquaintance’s follower list as a starting point for collecting data to be used in the study. Users were selected based on their geographical location within the Comunitat Valenciana by selecting users who self-reported their location as “Valencia.” An analysis of sociolinguistic factors such as sex and education level was not conducted due to the difficulty of establishing membership in these categories with any degree of certainty. Unlike social media sites such as Facebook, there is no requirement on Twitter to identify one’s gender on their profile. The only way to determine an informant’s gender would be to rely on an optional self-reporting of it. The same issues are found with regard to establishing the educational background of the informants. However, in terms to age, one might appeal to general trends on Twitter to establish that most informants are probably under the age of fifty, although this cannot be established with full certainty.

The Twitter corpus contains 106,719 words and 8,432 tweets for analysis. Twitter posts, or tweets, are limited to 140 characters; the user cannot exceed this character limit. The data was collected during the period of June 2014 to June 2015 from sixty different users. The size of the Twitter corpus was similar to the conversational corpora with regard to number of informants, although the later’s word limit exceeds the Twitter corpus due to the 140-character limit found on Twitter. Some factors impeded the continuous collection of data throughout this period of time. Some users stopped posting or closed their accounts and/or changed the privacy settings on
their account. In this case, data was collected for the time that it was available and stored for later analysis. The Twitter corpus differs significantly from the conversational corpora in the sense that the bilingual proficiency of each informant is unknown. A test was designed to alleviate this concern (mentioned below). Additionally, the Twitter corpus is comprised of planned discourse, i.e. each informant exerted a certain amount of planning when composing each tweet. This characteristic of the Twitter corpus allows for a quantitative analysis to be conducted on mixed language phenomena collected from non-spontaneous data sources to test the applicability of Poplack’s (2015) models on non-spontaneous data.

How was the geographic location of each Twitter informant determined? Only profiles that had indicated a geographical location were chosen for this study and data was collected only from users who explicitly stated they were from Valencia. For the purposes of this study, only informants who state ‘Valencia’ ‘Comunitat Valenciana’ or similar variations or cities within the autonomous region were used for data collection. The majority of Twitter profiles selected as sources for data are from users from the following cities: Valencia, Alzira, Gandia, Castellón, Cheste and Ontinyent. The following map indicates their location within the region.

**Figure 3. Map of Valencia (province)**
The following map shows specific Twitter data localities in relation to the city of Valencia.
Figure 4. Map of Valencia and surrounding area (city)
Sixty profiles were chosen from which to collect data for analysis. The number of tweets published by each informant varies considerably. Some users post very infrequently, with less than 15 tweets in a year, during the period of data collection. Some users post very frequently with around 500 tweets published during the period of data collection. The code and number of tweets for each informant is found in Appendix B. The users’ handle (‘profile’) name has not been revealed for privacy reasons. Throughout the study the users’ anonymity has been maintained by replacing any mention of a user’s handle name with ‘user’ or with codes assigned to each informant.

The type of informant sought out for data collection was a non-public figure, as public figures or companies for personal or product advertisement often use Twitter and the more scripted linguistic register often found in these accounts would be more distant from natural output. Any profiles with the words ‘official’ or the like of any team or organization were not included.

4.4 Data analysis and coding

Instances of the studied language contact phenomena were identified in the conversational corpora by the researcher, as were instances of similar language phenomena in the Twitter data. Each instance was analyzed and coded as either an instance of borrowing, code switching/language mixing or polylinguaging. Instances of borrowings were further coded as either established or nonce borrowings, according to criteria established by Poplack (2012, 2015) and Sankoff et al. (1990). This study distinguishes CS as an oral phenomenon and language mixing as a written phenomenon, due to differences in the planning of mixed language insertions in written discourse. However, the syntactic and discourse categories used to discuss these
phenomena naturally overlap to a certain degree. Previous studies discussing the phenomenon of polylanguaging (Jørgensen et al. 2011) did not conduct a quantitative analysis of polylanguaging. Therefore, this study subjects the model of polylanguaging to a quantitative analysis, an essential aspect of both contact linguistic studies and analyses on CS and borrowing from a sociolinguistic background (Poplack 2015). This study applies parallel constituent categorization to the non-oral data (with minimal adaptation) as was used to analyze the oral data while also noting differences in terminology and concepts (language mixing and polylanguaging). Due to a potentially lesser or complete lack of proficiency in L2 of a person who polylanguages, the extent of its use will be generally more limited to single word insertions and fixed expressions behaving syntactically like a tag, but not as limited in scope as what one finds in oral discourse.

The issue of categorization of linguistic phenomena in online data is a challenging one due to the fact that theoretical models of CS and borrowing were formulated in the context of oral discourse and often are inadequate when discussing written text, particularly taking into consideration the aforementioned issues related to CMC research. The underlying assumptions of conventional theoretical frameworks dealing with CS and borrowing do not necessarily correspond to realities found in online communication. Sebba (2012) notes,

“Analyses of spoken code-switching, whether or not they are done in a conversation analysis framework, are nearly always sequential; in other words, they treat the conversational data as a sequence of utterances (or turns) in which there is a more-or-less linear unfolding of the conversation. Indeed, the notion of a 'switch' itself relies on this property, containing in itself the idea of a 'now' in Language A and an immediate 'next' in Language B” (6).

For this reason, Sebba (2012) uses the term ‘language mixing’ to refer to written CS. The terminology of language mixing is used in this study to highlight that this is a fundamentally different process from CS in spoken communication. The motivations for including material from other languages in social media posts are different from communication offline. This study
looks to various approaches to CS and other related phenomena to establish an approach for analyzing such phenomena on Twitter, and study takes into consideration the arguments made by Flores and Lewis (2016) about approaching contact language phenomena from an emergentist perspective, which prioritizes context. Jørgensen’s (2011) term *polylanguages* will be used in describing some phenomena both on Twitter and in the conversational corpora. This study does not accept all arguments on polylanguaging; it is, however, an applicable term to indicate multi-language use in cases where it is likely that a speaker or social media user is perhaps not communicatively proficient in a given language. In this way, *polylanguages* can be used to describe some of the instances of language use found, and in differentiation from other language mixing phenomena. The term ‘code switching’ will be considered as a possible option to describe Twitter data where it is established that a user has also posted in Valencian and has a demonstrated proficiency in Valencian.

To code a language phenomenon as a borrowing, the criterion of morphological adaptation will be used as the primary tool (Poplack 2012, 2015; Sankoff et al. 1990). The Nonce Borrowing Hypothesis suggests that phonological adaptation is not absolutely necessary in cases of either established or spontaneous nonce borrowings. Additionally, the issue of Valencian and Spanish as typologically similar languages and phonological particularities of Valencian as a Catalan variety make phonological adaptation less observable due to congruence between the two phonologies. Valencian only possesses two additional phonemes, not found in Spanish, [ʃ] and [ʒ]. Valencian lacks the complex vowel system of Catalan and more closely resembles the Spanish vowel system. Second, due to the nature of the data compiled in this study being textual and lacking phonetic transcriptions of the corpora data and lack of any phonetic representation of the Twitter data, phonological adaptation cannot be used as a reliable tool to analyze language
phomena in this study. An additional consideration must be made to distinguish Valencian borrowings from instances of CS into Valencian due to the similarities in morphology in both languages. In this case, sites of conflict were analyzed to establish a word as a borrowing or not (Blas Arroyo and Tricker 2000), that is to say that if a Valencian noun in a Spanish context coincides with other Spanish nouns, then it is potentially a borrowing and not an instance of CS and vice versa. In conjunction with the analysis of morphological adaptation, the criterion of frequency was used to code lexical items as borrowings rather than CS due to evidence of their diffusion in the Spanish variety in the community (Poplack 2012; Sankoff et al. 1990). In this study this was accomplished by comparing a single word to determine if it occurred in the speech of more than one informant and across data sources. A lone insertion occurring only once in the data and meeting the criterion of morphological adaptation was coded as a nonce borrowing. Those found in more than one informant’s discourse were coded as established borrowings.

Lexical items not coded as borrowings, either established or nonce, are coded as either CS/language mixing or polylanguaging, depending on the observed bilingual proficiency of the informant who produced each instance. When analyzing the conversational corpora, the information provided by the compilers of the corpus was used to identify those speakers who were coded as proficient in Valencian. The exception was the Val.Es.Co corpus, which does not provide this information, in which case a similar test was conducted as for the Twitter analysis to determine if an informant produced syntactically complex material in Valencian to establish a degree of proficiency. The test mentioned above on the Twitter data was used to establish user proficiency in order to distinguish language mixing from polylanguaging.

Each instance of CS/language mixing was coded as either intersentential, intrasentential or a tag in each data source. The instances of polylanguaging found in this data are limited to
cases of hashtag use on Twitter. These are coded separately as “hashtag switches” in the next chapter presenting the results of this study. The nature of online communication was taken into consideration when classifying inter- vs. intrasentential mixing by acknowledging the differences in how discourse is structured online and acknowledging the complexity of tools such as hashtags which often overlap with intersentential mixing.

Additionally, when coding data from the Twitter corpus, the considerations of Jurgens et al. (2014) were taken into consideration to accurately code authentic instances of language mixing. First, there must be sufficient text to determine the message’s matrix language. This means that a one-word tweet or one with only one phrase would not be coded as a case of language mixing. I do not consider the alternation between tweet postings to be analogous to turn taking in a conversational setting. Second, it must be found in an act of communication and not as part of a spam-like or nonsensical message. This excludes coding instances of cultural references, product names, or other non-authentic messages as instances of language mixing.

Data analyzing CS and borrowings should consist of bilingual corpora, a sample of “spontaneous production of systematically sampled bilingual speakers resident in a well-defined bilingual speech community” (Poplack 2015: 921). This was accomplished in this study by analyzing data collected from the conversational corpora, representing a sample of bilingual corpora from two of the metropolitan centers of the Valencian Community and in the Valencian-speaking zone of the region. Information on the informants from the CSCDLP and PRESEEA corpora allowed for the bilingual proficiency of speakers to easily be established, using the information provided in the corpus itself. However, the Val.Es.Co corpus does not present information on informants who participated in the corpus. The representation of data from the corpora also allows an analysis of both completely spontaneous production (Val.Es.Co) and
semi-spontaneous production (PRESEEA and CSCDLP). The Twitter data obviously presents a different class of data. The analysis of data collected in a non-spontaneous source, Twitter, allows for the application of Poplack’s (2015) model of CS to a written discourse. To establish user proficiency on Twitter, a test was devised to measure evidence of proficiency via the inclusion of tweets in Valencian. Therefore, a tweet like the following was considered to be indicative of higher proficiency in Valencian.

Perfavor quanta hipocresia, molt de ficar els tituls d’Espanya i ara RT una foto aon posa q tornme la Copa Del Mundial 2010, esteu idiotes?

The following table shows how users incorporate Valencian and to what degree. Instances of language phenomena found in those profiles with posts in Valencian are more likely cases of language mixing, while those found in profiles with minimal or no posting in Valencian are more likely polylanguaging.

**Table 2. Language use per individual user**

<table>
<thead>
<tr>
<th>Language use</th>
<th>Number of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users who tweet in Valencian</td>
<td>13</td>
</tr>
<tr>
<td>Users who tweet minimally in Valencian</td>
<td>8</td>
</tr>
<tr>
<td>Users who tag in Valencian</td>
<td>7</td>
</tr>
<tr>
<td>Users who mix in Valencian but do not post in Valencian</td>
<td>19</td>
</tr>
<tr>
<td>Users who post exclusively in Spanish</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
</tr>
</tbody>
</table>

**4.5 Conclusion**

This chapter has presented the data sources analyzed in this study. Three conversational corpora were consulted: (1) the *Corpus sociolingüístico de Castellón de la Plana y su área*

Additionally, a corpus of data was collected from Valencian Twitter profiles to constitute a corpus of online data for comparison with the offline data collected from conversational corpora. A variety of conversational corpora were analyzed with the dual function of strengthening the comparison made with online data and also due to the fact that the conversational corpora were collected using different data collection methods, which reflect different modes of communication. This chapter presented the methods used to analyze the language phenomena found in each data source. Each instance of the language phenomena was coded as either an instance of borrowing, either established or nonce, code switching/language mixing or polylanguaging. To code a language phenomenon as a borrowing, the criterion of morphological adaptation will be used as the primary tool to code borrowings (Poplack 2015). Additionally, the criterion of frequency will be used to code lexical items as borrowings rather than CS (Poplack 2012; Sankoff et al. 1990). Special consideration was made to analyze morphological conflict sites to analyze and code Valencian insertions (Blas Arroyo and Tricker 2000). Lexical items not coded as borrowings were coded as either CS/language mixing or polylanguaging depending on the proficiency of the informant. Informant proficiency was established by informant data in conversational corpora and by considering users who include tweets in Valencian as evidence of proficiency on Twitter. Each instance of CS/language mixing was coded as either intersentential, intrasentential or a tag switch. Additionally, single and multiword insertions were coded and quantified according to appearance in each source.
CHAPTER FIVE
RESULTS AND DISCUSSION

5.1 Introduction

This chapter begins with a quantitative analysis of the data on CS and borrowings found in each conversational corpora used in this study applying the criteria suggested by Poplack (2015), which include the following: using the criterion of morphological adaptation to identify borrowings and distinguishing established borrowings from nonce borrowings (Poplack & Sankoff 1988; Sankoff et al. 1990) together with the criterion of frequency to code lexical items as borrowings or nonce borrowings (Poplack 2012). Poplack’s (2015) classification of CS into tag switches, intersentential switches and intrasentential switches is also applied to the data collected from conversational corpora. Blas Arroyo and Tricker’s (2000) considerations for analyzing Valencian and Spanish lone nouns were applied in this research by analyzing morphological conflict sites in coding lexical items.

I will then move to discuss the data collected from Twitter applying the same criteria to the greatest degree possible as well as the general ideas noted in Chapter 4 in the section on Methods as they pertain to handling instances of polylanguaging and the categorization of hashtags. The principle researchers here whose model I will follow include the following. Androtsopoulos (2008, 2011, 2013) provides much of the theoretical background as to how to apply research on CS from studies on conversation to online social media. The concept of polylanguaging (Jørgensen et al. 2011) is used in the analysis of Twitter data and quantified using methods for quantifying CS phenomena. This study applies Poplack’s (1980, 1981, 1990, 2015) model of CS and the Nonce Borrowing Hypothesis (Poplack 2012; Sankoff et al. 1990) to online data and to written discourse in that form.
Each instance of both CS and lexical borrowings is quantified in terms of absolute and relative frequencies for each data source. The types of syntactic constituents switched are also quantified. In terms of CS, the frequencies of intersentential, intrasentential and tag switches will be quantified in relation to each data source. Subsequently, a more detailed analysis of CS and borrowings will follow and then be supplemented by considering several factors that could influence the appearance of these phenomena. Additionally, when analyzing the Twitter data, the aforementioned issues with regard to language mixing and polylanguaging will be discussed.

The frequencies of occurrence of CS and borrowing will be compared across informants to determine if any informants had outlying frequencies that might skew the results. A few issues will be explored as unique factors in the use of CS such as the level of complexity of switches by looking at the frequencies of lone vs. multiword insertions. The issue of hashtags and their use will be discussed in detail as to their relationship to the rest of the tweet. The potential factor of linguistic economy will be considered as a motivation for CS online by taking into consideration the technological limitations set on the user by Twitter’s 140-character tweet limit. In other studies, linguistic economy has been used to explore the incorporation of borrowings. Separately, the use of English will be considered, which offers the majority of CMC polylanguaging, with regard to what advantages it offers the user.

5.2 PRESEEA

The following table presents the findings from the PRESEEA data on the absolute and relative frequency of occurrences of CS and borrowings vis-a-vis type of syntactic constituent. The PRESEEA data supports the Nonce Borrowing Hypothesis (Poplack 2012; Sankoff et al. 1990) in the relatively larger number of nonce borrowings, mostly lone nouns, which are generally the
most commonly found type of switch (Blas Arroyo 1998), however, the data from the PRESEEA results is relatively infrequent in comparison with other contact situations. With regard to CS, a few cases of noun phrases were coded as CS instead of borrowings. There were also some cases where a verb or verb phrase was more likely a CS than borrowing. The corpus also contains a number of cases of the use of CS for reported speech and in intersentential clauses in a roughly equivalent frequency. The PRESEEA corpus does not show any cases of English borrowings, however, it does show a large number of established Valencian borrowings.

**Table 3.** CS and Borrowing from Valencian into Valencian Spanish by constituent in PRESEEA

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>8</td>
<td>19.04 %</td>
</tr>
<tr>
<td>Verb</td>
<td>4</td>
<td>9.52 %</td>
</tr>
<tr>
<td>Adjective</td>
<td>2</td>
<td>4.76 %</td>
</tr>
<tr>
<td>Preposition</td>
<td>3</td>
<td>7.14 %</td>
</tr>
<tr>
<td>Discourse markers</td>
<td>2</td>
<td>4.76 %</td>
</tr>
<tr>
<td>Clause</td>
<td>10</td>
<td>23.8 %</td>
</tr>
<tr>
<td>Reported speech</td>
<td>15</td>
<td>35.71 %</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Borrowing</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonce NP</td>
<td>31</td>
<td>40.25 %</td>
</tr>
<tr>
<td>Nonce VP</td>
<td>4</td>
<td>5.19 %</td>
</tr>
<tr>
<td>Nonce Adj</td>
<td>5</td>
<td>6.49 %</td>
</tr>
<tr>
<td>Discourse markers</td>
<td>5</td>
<td>6.49 %</td>
</tr>
<tr>
<td>Established loanwords (NP)</td>
<td>32</td>
<td>41.55 %</td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

The results from the PRESEEA corpus show a low rate of inclusion of the language contact phenomena analyzed in this study. In a corpus of 575,000 words with 72 informants, the presence of nonce borrowings and CS is incredibly low. The use of established loanwords is also infrequent in the data.
5.2.2 Intersential CS

5.2.2.1 Nouns

The following show cases where nouns were judged to be cases of CS instead of nonce borrowings, due to the lack of morphological adaptation into Spanish. Both (1) and (2) show the use of Valencian morphological determination either in the lenited form associated with a pre-verbal forms in Valencian and Catalan.

1. De mi infancia / recuerdo / pues que / la forma de divertirnos y de jugar los niños y porque era en la calle / pues era tocar a la puerta de l’anella de la tía… (123.14 – el llamador) (F52VEM)
2. Me reulta un poco fuerte que maten a los toros// de todas maneras en els bous al carrer (371.97 – los toros por la calle) (M24VEM)

5.2.3 Intersentential CS

5.2.3.1 Reported speech

The PRESEEA corpus contains a number of cases of the use of CS for reported speech, where a speaker cites another person and uses the language used by that person. In this case, a speaker switches to Valencian to ‘copy’ the speech of another. However, as a generally noted characteristic of the PRESEEA corpus, the overall frequency is relatively infrequent in comparison with other contact situations.

3. Me dio un bofetón / en toda la cara // que yo cuando salí a las doce / de las monjitas / que me fui a casa // llego a casa y mi madre / ¿aixó qui te ho ha fet? / sor Francisca / mare / me ho ha fet sor Francisca… (124.41 – ¿eso quién te lo ha hecho? Sor Francisca, madre, me lo ha hecho sor Francisca) (F52VEM)
4. Unos amigos míos también tienen maquetas porque les han gustado y tal/ y claro y madre che és que ni han tantes/ que- que a vore si te desfás d’alguna a viatges (352.192 – che, es que hay tantas que a ver si te deshaces de algunas a veces) (M24VEM)

5.2.4 Lexical borrowing in PRESEEA
5.2.4.1 Established loanwords

The PRESEEA corpus presents a variety of single nouns that must be classified as either CS nonce borrowings or lexical borrowings. The criteria of morphological adaptation and frequency of usage are the two tools used to distinguish between established borrowings and instances of CS. In the case of Valencian nouns, morphological adaptation is often not possible to determine due to the many identical morphological features shared by Valencian and Spanish. In this case, frequency of usage is a more useful criterion to determine the presence of an established borrowing. Additionally, those Valencian lone items found in the discourse of Spanish monolinguals (as indicated by the information provided by the corpus) were then associated as borrowings if the first two tests did not suffice in coding an item.

5. Y luego mis mujeres estas de ahora pobretes pues… (436.419 – pobrecitas)
6. Y dije yo ya no vuelvo a examinarme y pobreta esta monja pues me convenció NADA (254.199 – pobrecita)

An example of a morphologically integrated borrowed noun would be in the following example, where we see present the Spanish masculine plural definite article los, which indicates complete morphological adaptation.

7. Quedan los abuelos/los yayos a los que vien a visitar los hijos… (259.263 – los abuelos)

There are, however, also other cases in the PRESEEA data of nouns of Valencian origin that should be classified as borrowings rather than instances of CS.

5.2.4.2 Nonce borrowed nouns

The PRESEEA corpus shows a number of nonce borrowings of Valencian NPs included in Spanish discourse. The following examples show the diversity in this regard, as there are feminine singular nouns, masculine plural forms, and masculine singular nouns. Some of these
nonce borrowings reflect phonological adaptation and some do not. However, they are behaving as Spanish nouns in this context.

8. Bueno/ pues/ estaba/ bueno los días que íbamos a ver los variatets/ luego el día del Cristo que pues en mi calle… (212.38 – variedades) (F29VEM)
9. Y también jugábamos al rogle/ que se llama aquí en Valencia (483.30 – al corro) (M50CEM)

The nonce borrowed items shown above show morphological adaptation into Spanish. The type of nonce borrowings found in the PRESEEA corpus show the use of nonce borrowings in cases to explain culturally specific material using the local term. In many cases, it seems like the informant using a local term intends to explain a particular local cultural issue to the interviewer.

5.2.4.2 Nonce borrowed verbs

The PRESEEA corpus shows several examples of Valencian verbs inserted into Spanish discourse, generally interpreted as nonce borrowings. As in the examples found above with NPs, the types of verbs found are also diverse. There are other cases in the corpus where Valencian verbs are used in a CS context which was determined by their lack of morphological adaptation.

10. eso que acostumbramos// te acomboyas a la mujer/ te vas al mercado… (493.364 – animar) (M50CEM)
11. No sé el nombre exactamente cómo se llaman/ pero antes habían/ unas máquinas que llevaban como una especie de manilla/ que sacaban el tubérculo/ lo espolsaban así (128.190 – sacudían) (F52VEM)
12. Sí es que me fique nerviosa es un grupo de eso y traen cantantes… (213.70 – es que me ponga nerviosa) (F29VEM)

5.2.4.3 Discourse markers

In data taken from the PRESEEA corpus, one finds Valencian discourse markers that should be considered borrowings rather than instances of CS. The discussion in 4.2.3, pointed out that oftentimes researchers are unsure how to classify discourse markers. . Here the fact that
these discourse markers, *a vore* and *ara* specifically, are found across data sources suggests that they are widely accepted and argues for a classification of them as borrowings. However, following Torres (2002), not all discourse markers can be coded as borrowings as such. The following example shows cases that are coded as borrowings. Spanish dominant speakers in the corpus use these. There are a few cases where discourse markers were coded as CS because they were only found in bilingual speakers’ discourse and appeared to not displace native options (*bé* for instance).

13. **A vore** el fin de semana pasado bueno el sábado por la mañana trabajé… (214.121 – *a ver*) (F29VEM)
14. También lo digo/ que para vivir no/ **ara**/ para pasar en verano (139.578 – *ahora*) (F52VEM)

### 5.3 CSCDLP

This section presents the data collected from the *Corpus sociolinguistico de Castellón de la Plana y su área metropolitana* (CSDCLP) with regard to the presence of CS and borrowing. The following table shows the findings from the CSDCP data with regard to the syntactic constituents. The CSCDLP data presents an overall increase in the presence of Valencian and English insertions in comparison to the PRESEEA data presented earlier. The prevalence of Valencian nonce borrowings is noteworthy as well as the a higher number of cases of English established borrowings. Valencian established borrowings and discourse markers are also found in bilingual and monolingual discourse. The CSCDLP corpus also shows the use of intersentential switching and the use of CS for reported speech. In comparison with PRESEEA, the CSCDLP data shows that CS is used mostly for reported speech and in intersentential switching through switching at clausal boundaries.
Table 4. CS and Borrowing from Valencian into Valencian Spanish by syntactic constituents in CSDCLP

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>1</td>
<td>2.12 %</td>
</tr>
<tr>
<td>Verb</td>
<td>2</td>
<td>4.25 %</td>
</tr>
<tr>
<td>Preposition</td>
<td>1</td>
<td>2.12 %</td>
</tr>
<tr>
<td>Adverb</td>
<td>5</td>
<td>10.63 %</td>
</tr>
<tr>
<td>Clause</td>
<td>18</td>
<td>34.04 %</td>
</tr>
<tr>
<td>Reported speech</td>
<td>20</td>
<td>42.55 %</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borrowing Nonce NP</td>
<td>35</td>
<td>30.43 %</td>
</tr>
<tr>
<td>Borrowing Nonce VP</td>
<td>11</td>
<td>9.56 %</td>
</tr>
<tr>
<td>Borrowing Nonce Adj</td>
<td>4</td>
<td>3.47 %</td>
</tr>
<tr>
<td>Borrowing Discourse markers</td>
<td>21</td>
<td>18.26 %</td>
</tr>
<tr>
<td>Borrowing Established loanwords (NP)</td>
<td>44</td>
<td>38.26 %</td>
</tr>
<tr>
<td>Total</td>
<td>115</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

While the CSCDLP presents a higher number of cases of CS and borrowing than that found in the PRESEEA results, the number of cases is still relatively infrequent in comparison to other analyses of CS and borrowing in contact situations.

5.3.1. Intrasentential CS

5.3.1.1 Nouns

Instances of single noun switches are rare occurrences in spontaneous data (Poplack 2015). One example was coded as a case of an intrasentential noun switch. This was due to the lack of morphological adaptation by using the Valencian plural formation allowing for the formation of a consonant cluster as found below.

15. Un tío mío y en fin/ bueno/ la fábrica era de carrocerías/ hacían tráilers/ camiones y nosotros estábamos trabajando/ pues estábamos ampliando la fábrica y: nos dejaban un torito para para: nuestro uso para si queríamos acercarnos palets: no sè de
ladrillos: de cemento de- si nos molestaba algo nos lo quitábamos// o sea/ estaba: para nosotros (Entrevista 29 – HE1SecVV)

5.3.1.2 Adjectives

In CSDCLP there are a few cases of the use of an intrasentential switch with an adjective. This was coded as CS for the same reasons as listed above with the intrasential noun switch due to lack of morphological adaptation.

16. La tía V. cuando son las tres “ahora vamos a descansar que estamos cansats” (Entrevista 21 – HE3PVV)

5.3.1.3 Adverb phrases

The data from CSDCP also has instances of AdvPs appearing as cases of CS.

17. Hombre/ la alta: (e)l agua de Cortes famosa/ que se anuncia/ y/ más (= a més = además)/ más que se anuncia (Entrevista 13 HE2PVV/C)
18. Cuando ya me pegué dije “uf! Menos mal/ no ha pasado(nada)/ ni le ha pega(d) a nadie ni me ha pega(d)o nadie” (Entrevista 45 – HE3C/VC)

5.3.2. Intersentential switches

The CSDCLP seems to offer the best examples of conversational CS used in a natural sense. In the following examples, we see a larger discourse where different triggers function within the discourse to motivate a switch.

19. ¿tú qué recuerdas con más cariño de tus años en la escuela? ¿qué recuerdas con más agrado de aquellos años? Quan anaes a escola?
   Tu anaes a les monges/ pero yo toda la letre que sé me la va ensenyar un mestre que estae en una mésia allì en el pueblo que decían A. y me van portar allí tot el curs sense estar a casa/ bueno/ lo menos tres anys o quatre/ allì/ y aquel estae castigat per roig/ y / estaba castiga(d)o y era de Alicante/ umm/ a ver S. de apellido/ ara no me’n recorde (Entrevista 22 – HE3PVV)
20. Lo que están diciendo es que/ _jo a lo millor per exemple no em donaràn beca/_ entonces _lo que pasara un poc es que cada_ Universidad hará un poco la selección hará como unas pruebas/ _per eixemple/_ yo estudiante de magistério de educación física/ _vale/_ e: para entrar aquí tienes que tener de media un ocho _I a lo millor la de_ Valencia es un cinco entonces qué pasa/ se darán también mas ayudas a unas universidades… pero soy de la Universidad de- puf! _I te diuen/_ pues mira/ _ixe titul que tens tu de metge de la universitat de pluf no val pa res i a lo millor dius/_ no no/ es que yo estoy licenciado en la Universidad de Stanford “I això es com si fores:// per qué? Perquè eixa universitat de Stanford demane unes mitges molt altes/ cobren molt per a paga- per a que puges estudiar allí/ entonces ¿qui va allí? _La gent de paper/_ I se fan universitats (Entrevista 22 – HE3PVV)

21. Vamos a ver/ hoy en día hay universidades privadas// J. va a lo de Universidad privada/ hace lo de INEF pero en una Universidad privada// entonces/ siempre habrá gente que tendrá dinero o no/ como los colegios/ “yo quiero que mi hija vaya a las carmelitas”/ o “_jo vull que la meva filla vaja al col-legi _son col-legis de/_ pero bueno/ por otra parte está la universidad Jaume I que es pública/ que tú puedes tener unos estudios con una calidad tal o la universidad que va J./ que bueno yo en el meu potencial economic <= tinc igual d’oportunitats que altra gent/ que de l’altra manera nomes té oportunitat eixei tipo de gent/ pero bueno/ no vamos a arreglar aquí el mundo (Entrevista 22 – HE3PVV)

22. Hombre: las fallas por supuesto/ _sóc valenciana filla/_ la traca/ _sóc traquera/_ festera/ y/ aunque no soy futbolera/ pues también me gusta saber que el Valencia gana (Entrevista 43 – ME3SecCC)

These examples reflect Blas Arroyo’s (1998) characterization of CS in Valencia as, “una alternancia de lenguas rápida, no balizada y principalmente intraoracional” (132). However, the overall rate of CS found in the CSCDLP is very infrequent.

5.3.2.1 Reported speech

Similar to what we saw in the PRESEEA data, CSDCP informants often use CS for reported speech.

23. Sí: que era- a ver/ yo es que me acuerdo de un poquito/ umm (canteando) “_ja has matat el gall/_ gall I gallina que encontramos manda el rey que lo matemos/ tris/ tras a pedraes morirás/ tris/ tras a pedraes morirás” ese es el himno (Entrevista 10 – ME2PVV)

24. Había un hombre que/// se llamaba el señor- el señor A. que nos decía “_ja vos agarraré jo/_ sin vergüenses/ endespués de que m(e) haveu- me haveu robat/ me
dieu que lo que vos haveu endut/ no se qué!”/ aixina// y nos ponía: ¡ay! Nos ponía como trapos (Entrevista 12 ME2PCC)

25. Entonces mi madre le dice “mire don A. esta xiqueta está tot el dia plorant/ yo no sé que le pasa” (Entrevista 18 – ME3PVV)

The intersentential switches found in the CSDCP are motivated by conversational cues between interviewer and interviewee especially when the latter is relating a personal story. In these cases, the speaker will use CS in reported speech to bring the original words of another speaker into the story.

5.3.3 Lexical borrowings in the CSDCLP data

The Valencian exhortative i avant appears as an instance of a borrowing in the CSDCP:

26. Y lo que es para uno pos se ve que está escrito que nos tiene que pasar/ i avant! Y ya te digo – y: porque: yo pues también gracias a Dios he tenido un carácter que: que me (he) sabido reponer a todas las circunstancias (Entrevista 18 – ME3PVV)

5.3.3.1 Nonce borrowings

Nonce borrowing of nouns is commonly found in the CSCDLP. The following examples show cases where a nonce borrowing was included.

27. Entonces por la parte esta de la calle C. pues ahí pusimos: la par·eta (= Val. Paradeta, Cast. Tienda) y ahí/ pues claro/ to(d)os los niños compraban allí (Entrevista 12 – ME2PCC)

28. Y tú preferías irte con el noviete: antes de: hacer tus: cosas/ y- y me dejaba los estudios y me iba corriendo con mi noviete / que tenía mi vespino entonces/ él tenía un vespino (Entrevista 13 – HE2PVV/C)

In the CSDCP data there are a number of cases of nonce borrowing verbs. The following examples show the use of a Valencian verb form in Spanish discourse.
29. Yo no creo/ yo no creo/ no/ yo en esas cosas soc (soy) poco cree (Entrevista 9 – ME2VC)
30. Entonces/ sí que tinnimos (tuvimos) el permiso del alcalde del pueblo (Entrevista 13 – HE2PVV/C)
31. Ya lo creo/ que volvería a estudiar/ no es que vuldria (quisiera) estudiar/ es que vuldria estudiar lo que aquello que/ que/ a mi siempre me ha gusta(d)o (Entrevista 14 – HE2PVC)
32. Mira si xarro (Entrevista 36 – ME2SecVV)
33. Es que no sé qué día va ser (fue) (Entrevista 21 – HE3PVV)
34. Mira allá cuando teníamos/ cuando teníamos allá donde va vindre don J./ el medico/ aquí (Entrevista 22 – HE3PVV)

The last two examples show the use of the unique periphrastic preterite, a feature that distinguishes Catalan from Spanish.

5.3.3.1 Discourse markers

Below are some examples of CS with Valencian discourse markers used in Spanish discourse. As in the PRESEEA corpus, these discourse markers are likely Valencian borrowings into Valencian Spanish.

35. “Home/ sí/sí/ alguna habrá pero/ no sé a (lo) mejor son típicas de esas que a lo mejor vas toda acelerada/ creyéndote que una persona es tal y no es/ te giras y dices “madre/ por Dios/ que: que que-“ porque a lo mejor vas chillándole de medio camino” (Entrevista 3 – ME1PCC)
36. ara a nivel de: de loterías o de suertes de: de cosas/ pos sí que una vez tuve suerte/ umm en la Caja Rural de donde te pasan la: la nómina me tocó un viaje (Entrevista 10 – ME2PVV)
37. no le gustaba ni que me pintara? Cada vez que me/ me- ¡y ya ves tú! ¿tú me ves ara lo que yo me pinto? (Entrevista 13 – HE2PVV/C)
38. tampoco no había nada bueno pues” ¡socorro socorro socorro! Llamar por teléfono a ver quién tiene un teléfono! A vore <-> a los bomberos socorro socorro! ¡que me se quema casa!” (Entrevista 18 – ME3PVV)
39. estoy de acuerdo per aixó/ perque (Entrevista 40 – HE2SecVV)
5.4 Val.Es.Co

The following table shows the data on code switching and borrowing vis-à-vis the syntactic constituency of the selected form as found in the Val.Es.Co corpus. The Val.Es.Co corpus presents results different from the other conversational corpora analyzed, likely due to differences in data collection: recall that the Val.Es.Co corpus is made up of secretly recorded conversations, instead of sociolinguistic interviews. The use of multiword CS is more frequent and is used in a higher frequency than in other conversational sources, accounting for 55% of the cases of CS. The use of CS for reported speech is also found in Val.Es.Co. It contains a very low number of Valencian discourse markers, used in Spanish. However, there are cases of Spanish discourse markers used in Valencian discourse.

**Table 5.** CS and linguistic borrowing from Valencian into Valencian Spanish per syntactic constituents in Val.Es.Co

<table>
<thead>
<tr>
<th>Intrasentential switching</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
<td><strong>Absolute frequency</strong></td>
<td><strong>Relative frequency</strong></td>
</tr>
<tr>
<td>Preposition</td>
<td>6</td>
<td>13.95 %</td>
</tr>
<tr>
<td>Clauses (intersentential)</td>
<td>24</td>
<td>55.81 %</td>
</tr>
<tr>
<td>Reported speech</td>
<td>13</td>
<td>30.23 %</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

**Lexical borrowing**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Established loanwords (nouns; Valencian &amp; English)</td>
<td>19</td>
<td>54.28 %</td>
</tr>
<tr>
<td>Nonce borrowing (NP)</td>
<td>1</td>
<td>2.85 %</td>
</tr>
<tr>
<td>Nonce borrowing (VP)</td>
<td>2</td>
<td>5.71 %</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>3</td>
<td>8.57 %</td>
</tr>
<tr>
<td>Adjective</td>
<td>10</td>
<td>28.57 %</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

A repeated pattern is observed in noting the relative infrequency of language contact phenomena when analyzing the Val.Es.Co results in comparison with the results from the other conversational corpora. In the Val.Es.Co corpus the use of CS and borrowing is even less
infrequent, due to the larger number of informants participating in the secretly recorded conversations.

5.4.1 Intrasentential code switching:

5.4.1.1 Verb

In Val. Es. Co there are a few cases of verb phrases that are more likely CS due to the lack of morphological adaptation, such as in the follow examples.

40. Eso está claro no que no m’apateix (Val. Es. Co 2.33 B:82)
41. ahora veremos o sea que no tenen dinés (Val. Es. Co 2.27 H:50)

5.4.2 Intersentential code switching:

5.4.2.1 Reported speech

The Val. Es. Co corpus has a number of quote switches of which the following are examples. CS for reported speech is found more in the conversational corpora and less in the Twitter data.

42. Seguro que llega a mi madre y le dice que diu la xica no sé què d’un cotxe (Val. Es. Co 2.11 A:19)
43. su madre el aviso a lo millor no se’n recordará que ha de vindre (Val. Es. Co 2.2 I:191)
44. y dijo ARA que tenim dos diputats en Europa anem a luchar per València TOMA deme una botellita de agua señora Rosario ara anem a luchar per València dic jo sí adeprén primer a parlar fill de puta (Val. Es. Co 2.32 F:129 – two instances of reported speech here)

5.4.3 Lexical borrowings:

5.4.3.1 Established borrowings and nonce borrowings
The Val.Es.Co corpus has nonce borrowings for nouns like the other corpora as the following examples show.

45. este es mi paquet (Val.Es.Co 2.3 B:86)
46. ai pobra dona / ya basta que te entienda (Val.Es.Co 2.12 A:15)

The Val.Es.Co data shows relatively less nonce borrowings from Valencian into Spanish, in fact, it is surprising how infrequently they occur. The Val.Es.Co data also shows relatively infrequent use of English borrowings, much like PRESEEA, but dissimilar to CSCDLG.

5.4.3.2 Discourse markers

Val.Es.Co also has discourse markers, which appear to be borrowings applying the criteria previously mentioned. The following examples show both the inclusion of Valencian discourse markers into Spanish discourse and the inclusion of Spanish discourse markers into Valencian discourse.

47. Está roncando ay mare mía (Val.Es.Co 2.33 B:46)
48. ¿entonces vindrás di- dilluns per la nit? (Val.Es.Co 2.2 I:32)
49. está roncando, ay mare mía (Val.Es.Co 2.33 B:46)

5.5 Twitter: Basic considerations

The presence of language mixing is at about the same frequency as that found of one composing tweets exclusively in Valencian. The table below presents which languages informants mix on Twitter. The table shows instances of alternation of Spanish, Valencian, and English in a tweet. The following tweet presents a case where a subordinate clause in Spanish alternates with Valencian in a tweet.
50. Els comentaristes de Telecirco, que Costa nos aportará cosas a la selección… si… mos portará a casa jajaja

The following table shows where alternation occurs into Spanish, Valencian, and English. It does not present the language mixed from, i.e. in the example above there is alternation from Valencian to Spanish at the clausal boundary. However, in the table below, this example would fall under the category ‘Tweets with alternation to Castilian.’

Table 6. Embedded languages on Twitter

<table>
<thead>
<tr>
<th></th>
<th>Absolute frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets with alternation to Castilian</td>
<td>38</td>
<td>14.17 %</td>
</tr>
<tr>
<td>Tweets with alternation to Valencian</td>
<td>135</td>
<td>50.37%</td>
</tr>
<tr>
<td>Tweets with alternation to English</td>
<td>95</td>
<td>35.44 %</td>
</tr>
<tr>
<td>Total</td>
<td>266</td>
<td>100.0 %</td>
</tr>
</tbody>
</table>

The table shows that users prefer switching into Valencian, with English following in second and lastly switching into Spanish. There is less switching into Spanish, as Spanish is generally the preferred code of users. It is not surprising that users do not mix English with more frequency on Twitter as mixing, like code switching, implies a higher level of speaker proficiency in the embedded language than other forms of polylanguaging, although the use of English would serve to expand the audience of a user’s tweets.

Table 3 in Chapter 4 demonstrated that 22 users in the Twitter corpus post in Valencian with varying frequencies. In these cases, a more convincing argument can be made to have the language mixing that appears in their tweets. With instances of language mixing in other profiles, the argument of informant proficiency and of an instance of language mixing as potentially being one of CS is more problematic and additional factors need to be considered.
Instances of language mixing in intersentential contexts are likely ones of CS as well as where it appears in subordinate clauses. In instances where English is the embedded language it is less likely to be an instance of CS and these cannot be reliably classified as such. Applying a broader notion such as that of language mixing both captures a fuller range of contact phenomena as well as allows the researcher to avoid mislabeling informant output as instances of CS when there is insufficient information to do so. It additionally highlights the degree of importance pertaining to the context of the act of communication.

With regard to instances of English hashtags, these are cases of polylanguaging, in which case competency is separated from language use so that a speaker can acquire individual linguistic features from a language and incorporate those into their speech. The Twitter data also shows occurrences of polylanguaging into other languages, such as French, Italian and Portuguese. The arguably low likelihood of user proficiency in those languages suggests that these should be classified as instances of polylanguaging rather than language mixing. Although the term polylanguaging reflects a more theoretical conceptualization dealing with how a speaker can use elements from other languages while lacking advanced proficiency in a language, cases of polylanguaging were collapsed in with language mixing in the instance of the above table to specifically focus on degrees of appearance of the different constituency categories found ‘mixed’ in the Twitter data.

The tables below document this complex linguistic situation. Taking the previous discussion into consideration:

1. Instances of tag switching on Twitter include mostly Valencian tags but a few cases of English tags.
2. Greetings and farewells are separated from tags in the following table due to their differing pragmatic function.

3. Instances of intersentential switching are likely cases of CS and are labeled as ‘switching’ below.

4. Hashtag mixing was separated from intersential switching due to the difference in their behavior and incorporation. These are mostly English insertions and are treated as polylanguaging.

5. Intratextential switching occurs minimally in the data but represents the behavior of a more proficient bilingual.

The total is different from the numbers in the previous tables because this focuses on the total number of instances of CS while the tables above present the number of tweets with CS and many tweets contain more than one instance of CS.

Table 7. Tags, intersentential and intratextential mixing on Twitter

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags</td>
<td>99</td>
<td>25.84 %</td>
</tr>
<tr>
<td>Greetings and farewells</td>
<td>81</td>
<td>21.14 %</td>
</tr>
<tr>
<td>Intersentential switching</td>
<td>52</td>
<td>13.57 %</td>
</tr>
<tr>
<td>Hashtag polylanguaging</td>
<td>52</td>
<td>13.57 %</td>
</tr>
<tr>
<td>Intratextential switching</td>
<td>5</td>
<td>1.3 %</td>
</tr>
<tr>
<td>Borrowings (nonce and established)</td>
<td>89</td>
<td>23.23 %</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>100.00</td>
</tr>
</tbody>
</table>

5.5.1. Hashtags

Hashtags were created by Twitter to assign a topic to a Tweet. So, for example, Tweets that contain #WorldCup are about just that. One clicks on a hashtag to see Tweets related to a topic. In other words, the hashtag uses a system called tagging in order to “integrate a tweet into
a conversation on a given topic” (Wikström 2014:132). In a hashtag, “key concepts are distilled into short strings that are then attached to a document, image, or resource to facilitate retrieval” (Huang et al. 2010:1). They have a lifecycle, which begins with their adoption by a digital community after their creation by a specific user and can grow in popularity but, in the end, all hashtags suffer from abandonment by the users in favor of another hashtag. Therefore, hashtags have the function of connecting ideas between different users about a shared theme.

Twitter users have expanded the use of hashtags to use them in other functions beyond their original context. Cunha et al. (2011) note the potential uses of hashtags on Twitter beyond that of connecting posts on the same theme:

“They can be used not only to add context and metadata to the posts, but also for promotion and publicity. By simply adding a hash symbol (#) before a string of letters, numerical digits or underscore signs (_), it is possible to tag a message, helping other users to find tweets that have a common topic. Hashtags allow users to create communities of people interested in the same topic by making it easier for them to share information” (58).

As a point of clarification, the concept of hashtags (unique to Twitter) should not be confused with tag or tag switches, which are insertions of emblematic material in a category separate from intersentential and intrasentential CS (cf. Poplack 2015). With the exception of a few instances of an intrasentential hashtag clearly inserted into the user’s tweet (noted in examples), hashtags can best be understood as instances of either intersentential CS (as they are often classified in CMC research) or of tag switches depending on the specific case, as was applied here in both the tables and examples.

Wikström (2014) studies the uses of hashtags in relation to a variety of communicative and pragmatic functions. He enumerates the following uses of hashtags in online discourse (1) using them in games, (2) use as meta-commentary about the theme of a tweet, (3) for elaboration, (4) to add emotion, (5) to emphasize, (6) for comedic purposes, (7) to make cultural
Hashtags are characterized according to Wikström by their unoriginality vis-à-vis the user. Wikström also notes the use of hashtags as a substitute for features that Twitter lacks, such as bold or italics in a Word processor as well as for symbols that Twitter does have such as asterisks or parentheses.

Jurgens et al. (2014) studied the use of CS in hashtags. According to them, users are motivated to include hashtags in other languages to either increase their potential audience or to claim membership in the online multilingual community. They note that when there is an increase in the amount of bilingualism, the frequency of hashtags in other languages decreases, while the overall rate of CS increases. The dual purpose of hashtags is a way of marking particular content and also as a method of community formation through the discussion of a particular theme.

Regarding the hashtags found in my Twitter corpus, the following observations can be made. Hashtags are generally material independent from the rest of the material in the tweet; a tweet can be composed of only a hashtag which presents a complete thought:

51. #LoveSexAndRockAndRoll
The following example shows a hashtag related to previously mentioned material but separate from the main content:

52. Bueno vaaa, que las maletas no se hacen solas…#PlayitaAlláVoy
Another example shows a hashtag functioning as an addition to the main content:

53. A dormir, que estic feta pols! #BonaNit
In (56) hashtags are syntactically integrated into the main content appearing both at the front and at the end of a tweet. The first tweet offers commentary on what follows and the second is part of
the sentence, but because it refers to a film it has been highlighted with a hashtag. (57) shows another case where the hashtag is syntactically integrated into the tweet.

54. #TwitterOFF q [sic] me voy a ver #CiegaACitas ☺
55. #MeDanGanasDe estar en la playita tomando sol el sol y bebiendo

An analysis of the hashtags above suggests that a hashtag can be (1) classified as something similar to a tag switch, (2) somewhat like an intersentential switch, or (3) integrated (similar to an intrasentential switch). Hashtags then do not fit into one category but must be examined in context, weighing both syntactic and semantic considerations.

5.5.2. Syntactic Constituency of Mixed Elements

Table 9 indicates the syntactic constituents mixed. It presents the syntactic constituency of inserted elements, those forms different from the main language used in the tweet.

Table 8. Constituents ‘mixed’ on Twitter

<table>
<thead>
<tr>
<th>Intrasentential mixing</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preposition phrase</td>
<td>4</td>
<td>3.36 %</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>1</td>
<td>0.84 %</td>
</tr>
<tr>
<td>Clause (intersentential)</td>
<td>57</td>
<td>47.89 %</td>
</tr>
<tr>
<td>Hashtag switches</td>
<td>52</td>
<td>47.89 %</td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>100.0 %</td>
</tr>
<tr>
<td><strong>Tags</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tags other than those below</td>
<td>95</td>
<td>52.77 %</td>
</tr>
<tr>
<td>Greetings and farewells</td>
<td>81</td>
<td>45.0 %</td>
</tr>
<tr>
<td>Happy birthday wishes</td>
<td>4</td>
<td>2.22 %</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>100.0 %</td>
</tr>
<tr>
<td><strong>Lexical borrowing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonce noun</td>
<td>19</td>
<td>21.34 %</td>
</tr>
<tr>
<td>Nonce verb</td>
<td>11</td>
<td>12.35 %</td>
</tr>
<tr>
<td>Nonce adjective</td>
<td>6</td>
<td>6.74 %</td>
</tr>
<tr>
<td>Established noun</td>
<td>36</td>
<td>40.44 %</td>
</tr>
<tr>
<td>Discourse markers</td>
<td>17</td>
<td>19.1 %</td>
</tr>
<tr>
<td>Total</td>
<td>89</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>
The Twitter data follows a similar pattern to other data sources in the use of nouns. However, the Twitter data contains a higher number of English established borrowings (discussed below in Sec. 5.5.6). The Twitter data shows a minimal occurrence of intrasentential mixing.

The issue of distinguishing between intersentential and intrasentential switching on Twitter must take into consideration the differences between how communication is organized and structured differently offline and online. An intersentential switch occurs at a clause or sentence boundary or at the change of turns in a conversational setting. Turn taking, which is integral to conversation, does not really occur on Twitter, which leaves the distinction between intersentential and intrasentential switching to be distinguished by looking solely at the clause or sentence boundary.

Punctuation is generally used as a tool to classify intersentential switches versus intrasentential switches and Twitter users rely only sometimes on standard punctuation to organize discourse. Some authors have even noted that some Twitter users would prefer a more formal language on Twitter (Hu, et al. 2013). The principal punctuation mark that is lost or discouraged on Twitter is the period. Gunraj et al. (2016) suggest that using a period when text messaging is perceived of as insincere, overly formal, distant; the same perception appears to exist online with Twitter. However, on Twitter an alternative system of symbols is used to highlight or mark off certain aspects of discourse and often serve as a proxy for tools available to speakers in other contexts. For instance, the use of all capitals can be used in conjunction or in lieu of an exclamation point, as a way to express emotion.

56. GRACIAS A TODOS LOS QUE OS HABEIS QUEDADO PARA DELEITAROS CON ESTOS MONSTRUOS. OS QUIERO #ULTRALIVE #Ultra2015 (F1A3)

Users often use ellipsis as a way of separating clauses.
57. Guetta no, mejor Armin o dormir... que ya no puedo más (F1A3)
58. Me voy a las 19:30 y no sé a qué hora volveré y no sabré que sets me voy a perder... no me gusta :/ (F1A3)

Hashtags can also be used as a form of punctuation, sometimes functioning to mark a separation between clauses.

59. Por hoy ya hay bastante #BonaNit (F2A4)

5.5.3 Intrasetential switches on Twitter

In contrast to intersentential code switching, intrasentential code switching refers to the use of more than one language within a phrase or sentence. The inclusion of intrasentential switching is minimal in the Twitter data.

5.5.4 Inter sentential switches on Twitter

Intersentential switching on Twitter refers to language mixing that occurs at the sentence or phrase level. In conversation this happens via pauses; on Twitter the use of orthography clarifies that a new sentence, phrase, or thought has begun. A Twitter user can also use other tools such as the hashtag (#) symbol.

60. Suena ACDC #i'mhappy (M3G3)
61. Pensar demasiado las cosas puede hacerte perder grandes momentos y oportunidades... now is the moment!!! (F4A2)
62. Hace un poco (demasiado) fresquiiito! #WelcometonorthernEngland (F5V1)
63. "@user: Te necesito como el aire para respirar.26 pic.twitter.com/TLp5r9mMgD” no te puc voler mes carinyo (M10V1)
64. Pérez de Vargas is the man!! Ahora queda rematar la remontada!! Vamossssss!!! (M19V24)
65. Comença el partit!! #Amunt Me ha fallado Nuno no poniendo a @user (F21V31)
66. Se acabó por hoy... bona nit y a descansar. (F4A2)
67. Siesta con tormentor..lo millor del mon!... http://youtu.be/HyJm2WtnnK0 (M25V33)
There are a number of cases that I am classifying as intersentential switches where users and their followers use an application, called ‘ask.fm’. In the application, users can ask each other questions. A user can type a question for specific friends or post questions anonymously. The question will show up on a page titled ‘Questions’ where the user has all of the questions asked by friends as well as generic questions provided by the website, such as ‘What are you doing?’ In many cases, the question is written in Spanish but the answer is provided in Valencian.

68. ¿Qué tipo de zapatos llevas más? — **Deportives, sempre en les meues Converse** \(^{\text{^^}}\) \(\text{http://ask.fm/a/bebl05je} \) (F2A4)
69. ¿Cuál ha sido la última bebida que has tomado? — **Aigua, que fa la vista clara! xD** \(\text{http://ask.fm/a/anjaf019} \) (F2A4)
70. Cuando tengas 90 años, ¿qué es lo que más te importará? — **Anarmen de festa jejejeje** \(\text{http://ask.fm/a/b1lqe3ej} \) (M10V1)

Many times an intersentential switch coincides with a hashtag.

71. Demà més faena ... **#QuieroRelax #BonaNit** (F2A4)
72. Canción Newcastle:// **https://www.youtube.com/watch?v=O-zpOMYRi0w ... #imsofancy** (F5V1)

The hashtag below however does not reference a well-known tag expression and seems to better reflect an intersentential switch.

73. Primer día superadoooooo! **#Englishteacher #Elvedat** (F5V1)

Sometimes there is a complete thought expressed in a tweet followed by a string of hashtags. I interpret these instances as intersentential switches because the hashtag, usually in English, is not really connected to the main content of the tweet. The literature allows for intersentential switches to occur at the phrase level. In some of the cases below, there is a string of NPs such as (74), (76) and (77), which contains an AdjP in a hashtag with an adverb assigned to it but not included in the hashtag.

The context of these hashtags could be interpreted as polylanguaging, although the context, particularly of (74) and (75), seems to indicate that they are used contextually; however
CMC researchers would consider these as instances of CS as the only known study on CS in hashtags (Jurgens et al. 2014) considers such instances to be ones of CS.

74. De piscineo #summer #benitatxell #pool (F2A4)
75. WOOW #LetItGo de @user & @user en el #N15 de #Maxima51Chart enhorabuena chicos! ;) Seguimos subiendoooo! very #happy (F2A4)
76. Te voy a echar de menos! #music #love (F2A4)
77. Ostiaaa! Que gran noche!!! #perfect #forever (F2A4)

5.5.5 Tag Switches and polylinguaging

Table 8 above includes ‘tags’ as a separate category even though the table intends to principally delineate the syntactic constituency of mixed elements. The reason is that tags are highly formulaic fixed expressions that may or may not overlap with a syntactic constituent. Tags can be NPs (Bon Dia, Bona Nit, etc.), VPs (Junts Tornem), or a type of exhortative interjection (Amunt) while still maintaining their membership in what seems as the highly significant ‘tag category’. Here they might also appear as hashtagged material. The low frequency of ‘happy birthday wishes’ in the data here is noteworthy as CMC researchers have noted that it as a popular vehicle for language mixing (Androutsopoulos 2013). If one folded the rather large number of instances of tags into the categories of ‘NP’ or other syntactic categories that would artificially inflate the numbers in those categories and hide the importance of ‘tags’ as an influential category in discussing the language mixing phenomenon.

Intersentential switching is more complicated syntactically than tag switching because the user must follow the rules of syntax. However, intersentential switching is less syntactically integrated and, therefore, less complex than intrasentential switching in which a user must fully integrate the two codes into one discourse. Intersentential switching in the Twitter data was judged to be a genuine occurrence of language mixing due to the syntactic complexity of the
material switched as well as the occurrence of intersentential switching in principally in profiles of users with tweets in Valencian. The question of how to accurately differentiate between intersentential and intrasentential switching online is often a challenge as it involves defining online sentences and clauses, the sites at which intersentential switching occurs.

Poplack (2015) suggests that tag switching does not affect the overall meaning of an utterance and that they can be deleted or moved with little semantic consequence. A tag switch is unlikely to violate grammatical rules. She relates the use of tag switches to the speaker’s level of proficiency in the L2, “the less knowledge the speaker has of the other language, and less structurally integrated into the discourse will be his switches into that language” (1980: 171). The low level of proficiency required to use tags conforms to the ideas postulated by Jørgensen et al. (2011) in their framework of polylanguaging, although polylanguaging does not posit a limit in terms of what a speaker or user can incorporate. This suggests that tags and polylanguaging are similar with regard to the degree of proficiency required to use them in discourse. However, the difference between these two linguistic phenomena has to do with the issue of the degree to which they are emblematic material. Poplack (1980) further describes the situation, showing the emblematic nature of tag switches.

“[T]hose who claim Spanish as ‘the language they feel most comfortable in,’” and also evidence positive attitudes toward its use and maintenance, tend to switch into English primarily through use of tags, frozen forms, and idiomatic expressions. Thus 82 percent of the English switches transcribed for Juan, a key Spanish-dominant figure on the block, were of this type: e.g., “right,” “you know.”” (171).

She describes a situation similar to what is found in the Twitter corpus. The high frequency of tag switches could indicate an overall low proficiency in Valencian by these users; however, it does not necessarily need to imply that. At the same time, a competent speaker can engage in tag switching and other forms of emblematic switching in conjunction with intrasentential switching,
which is seen in the Twitter data through the use of tags by those users who post in Valencian. The type of items used in polylanguaging as found in Jørgensen et al. (2011) are not emblematic in nature. This differentiates polylanguaging from tag switching by noting that polylanguaging is not necessarily emblematic in nature.

Lipski (2005) notes the pattern of tag switching among US Latinos, especially among those who are not proficient in Spanish, which is some ways parallels the type of tag switching found in the Valencian Twitter data. The presence of a similar type of pattern across various populations is interesting and notes the more universal nature of tag switching as an active way that speakers can incorporate language as a way of expressing identity and group membership.

The following table shows the frequency of each type of tag in the Twitter corpus, many of which appear within hashtags.

Table 9. Tags by function on Twitter

<table>
<thead>
<tr>
<th>Function</th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greetings and farewells</td>
<td>81</td>
<td>45.0 %</td>
</tr>
<tr>
<td>Exhortative function</td>
<td>85</td>
<td>47.22 %</td>
</tr>
<tr>
<td>Explitive function</td>
<td>5</td>
<td>5.0 %</td>
</tr>
<tr>
<td>Vocative function</td>
<td>5</td>
<td>5.0 %</td>
</tr>
<tr>
<td>Happy birthday wishes</td>
<td>4</td>
<td>2.22 %</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

The following table shows the frequency of each specific tag.

Table 10. Tags used on Twitter

<table>
<thead>
<tr>
<th>Tag</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amunt</td>
<td>67</td>
</tr>
<tr>
<td>Bona nit</td>
<td>59</td>
</tr>
<tr>
<td>Bon dia</td>
<td>16</td>
</tr>
<tr>
<td>Mare</td>
<td>8</td>
</tr>
<tr>
<td>Good night</td>
<td>5</td>
</tr>
<tr>
<td>Junts tornem</td>
<td>5</td>
</tr>
<tr>
<td>Crack</td>
<td>5</td>
</tr>
<tr>
<td>A fer la ma</td>
<td>4</td>
</tr>
</tbody>
</table>
There is a slight preference (55.68 %) for tags to appear without a hashtag, in which case, the purpose of attaching a tag switch to a hashtag would relate to the more general purpose of hashtags as a mechanism to connect a tweet with tweets with the same hashtag.

5.5.5.1 Tags with exhortative function

Tags in the Twitter corpus can be classified under a variety of different functional types. The most common type here is an *exhortative tag*. The Valencian word *amunt* appears in a number of contexts in the corpus either narrowly in a sports context to express team affiliation and support or more broadly as a way of expressing a Valencian cultural identity. The *Diccionari Normatiu Valencià* defines the word as: interj. “S'usa per a animar a pujar, o a començar o a prosseguir una faena o una empresa. Amunt València!” (Diccionari Normatiu Valencià), i.e. an exclamatory term used to encourage, begin or continue an activity. It usually is found unmodified, and can be included in a hashtag. Sometimes it is repeated or included as part of another expressions such as with salutations.

78. Siempre duermo con ella y por siempre #OrgullosoDeSerValencianista #AmuntValencia (F1A3)
Greetings and farewells

One of the more frequently found tags in the Twitter corpus is a greeting or farewell in a language other than that of the principal language of the tweet. For instance, if the principal language of a tweet is Spanish, the tendency would be to include a salutation in English or Valencian. Instances of salutations are found in all three languages, either as instances of polylanguaging or in monolingual tweets. The most common salutation found in the corpus is the farewell BonaNit in many variations, including with or without a hashtag; ‘bonanit’ ‘bona nit’ ‘BonDia’ and ‘bye bye’ or ‘goodnight’ are possible options for a salutation. In (1) and (4), the
salutation is repeated in Spanish and Valencian, perhaps serving to acknowledge both Spanish-speaking and Valencian-speaking followers.

The following are instances of salutations found in the Twitter data.

89. Buenisimas noches a todos, me voy con el mejor recuerdo, el set de Alesso hoy en el ultra miami #bonanit (F1A3)
90. Llegamos a casa, bona nit (F2A4)
91. Nos hemos levantado #BonDia (F2A4)
92. Adioos Messi, como tu falta... bye bye (F2A4)
93. Me encanta ver el cielo por la noche #relax #GoodNight :) (F2A4)
94. Tengo sitio en el parking con sombra!!!!! #goodmorning (F5V1)
95. Buenos dias!! Bon dia!! Me encanta vivir en el Mediterraneo. BENICARLO!! (M6C1)
96. "@Mister_CeS: Buenas noches a toda mi gente con SUAJ." @raquellauri99 good night baby (M10V10)
97. Cuando un sentimiento habla por si solo ya no hay nada que hacer, no habrá nada capaz de pararlo #Bonavesperada (F14V19)
98. Buenos dias!! Bon dia!! Me encanta vivir en el Mediterraneo. BENICARLO!! (M6C1)

5.5.5.3 Happy birthday wishes

Another type of tag is happy birthday wishes whereby a user will direct a happy birthday wish to a specific person (Androtsoupolous 2013). Often, this birthday salutation will be addressed to the referent in their native language indicating linguistic deference.

99. Hoy cumple 29 años @MichaelPhelps Felicidades!! #HappyBirthdayMichaelPhelps (F1A3)
100. Moltes felicitats pels 28 @user !! passa un bon dia #HappyBday :) (F1A3)
101. Un sentimiento no trates de entenderlo, felic aniversari @valenciacf (F18V28)

Related to the above examples are those of holiday greetings.

102. Primer chocolatito del calendario de adviento #merrychristmas (F5V1)

The differences between Facebook and Twitter can be seen in comparing the relative scarcity of happy birthday wishes in this data in comparison with studies such as Theodoropoulou (2015).
The length of data collection should account for some of discrepancies (1 year vs. 3 years of collection). However, beyond that, the Twitter data shows only 4 cases of birthday wishing. Additionally, three out of four are not triggered by a personal acquaintance but directed to sports celebrities or to sports teams. Twitter and Facebook differ in this respect as well. On Facebook, users are prompted by Facebook to wish their Friends happy birthday on the date of the birthday. Twitter does not include this prompting. Also, Twitter and Facebook are different mediums of social media interaction and should be treated separately, except in cases where their use overlaps, such as in wall posts on Facebook or tweets on Twitter.

5.5.5.4 Tags with an expilitive function

A number of Valencian and English expressions are included in a Castilian text, which is fixed or emblematic expressions. (104) shows an instance of an English expilitive in a Valencian text; the Valencian expression *a fer la ma* translates roughly into Castilian as *a tomar por el culo*.

103. A los madrístas que hoy animan tanto al Valencia, una cosita: Iros **a fer la ma**!!
(F16V23)  
104. Jo que volia anar-me'n contenta a dormir... i al final encara ha empatat Portugal... 2-2 #USAvsPOR #shit
105. **A fer la ma**!!! Corre la banda @user_

5.5.5.5 Motivations for Tag Switches and polylanguaging

The motivations and potential outcomes for tag use will be discussed. Androutsopoulos (2011) mentions the emblematic use of English phrases in CMC:

“Studies suggest patterns of ‘minimal bilingualism’… in which sets of English chunks and formulaic routines (including greetings and farewells, interjections and discourse organisers, requests, slogans, etc.) are inserted into the base national or majority language. Their choice is often indexical to the groups’ lifestyle orientations, including stylized representations of vernacular ‘Englishes’” (Androutsopoulos 2011: 11).
The relationship of CS to identity has been studied extensively. Zentella (1997) found that speakers included short code shifts into a larger discourse; this had the effect of “continually reasserting and recreating children’s dual New York-Puerto Rican identity” (120). In discussing the issue of the use of CS as “we-code” (Gumperz 1982), each bilingual context must be analyzed individually (Sebba and Wootton 1998). The use of Valencian in this data set does not have an exclusionary function, i.e. it is not meant to set group boundaries in the sense of excluding some from group membership. In some bilingual contexts this strict binary group membership via code choice might not be as binary as in other communities, especially in the case of Spanish regional identities. The overall “Spanish” identity might override the regional “Valencian” identity. However, the data does seem to indicate that users are performing identity via language practices. This might be a version of “acts of identity” (Le Page & Tabouret-Keller 1985), who note that linguistic behavior can have the effect of being an act of identity. Sebba and Wootton (1998) discuss that in order for language to have an identity function it must be used in talk. This data suggests that this function can also be written or *posted* online. Given that the majority of the cases of tag switching on Twitter are Valencian insertions (79 %), this discussion of the use of CS in identity performance is an important factor in determining the possible motivations for the inclusion of these tags in the Twitter data. Given that a majority of the users in the sample of data analyzed from Twitter include some form of mixing in their tweets, around 80 %, this seems to suggest that Androutsopoulos (2011) notion of minimal bilingualism should be extended in cases of linguistic areas with a regional language so that the regional language will be included in a similar way as a form of regional minimal bilingualism.

On Twitter, users are hypersensitive to issues of identity performance and rely on self-monitoring techniques to constantly construct and modify identity in their presentation of self
often as members of the ‘we’ group. This is due to the nature of Twitter as a social media site focused on the individual and not on interaction between users. In consideration of tags, their use in a repeated and highly formulaic way would indicate that their use by Valencian users is a way of constructing and performing an online identity. The effect of context collapse, discussed earlier, is a blurring of varied social relationships into one audience. Interactions on Twitter “reproduce existing ties and patterns in sociality, but also vary in conversationality depending on both actual and imagined audiences” (Papacharissi 2012: 1994). It can be assumed that the Valencian informants are aware that they are ‘speaking’ to a mixed audience that conflates many different context-sensitive offline relationships into one online one.

The data in this study leads to an interpretation of Twitter as a public domain in which users might not favor the use of Valencian due to the lack of an intimate context. This relates to the issue of context collapse because all of a user’s contacts are mixed into one setting. This disfavors the use of Valencian and reflects the common situation in a bilingual context where one language is dominant and the other subordinate that there is a preference for switching into the dominant from the subordinate language (Lipski 2005: 8).

Is there anything comparable in the other corpora data like the use of tag fixed expressions on Twitter? In the PRESEEA corpus, there were no tag switches such as the greetings or farewells found on Twitter in the sense of a fixed expression. The only comparable phenomenon is the use of discourse markers in terms of their more variable placement.

106. También lo digo/ que para vivir no/ ara/ para pasar en verano (139.578 – ahora)
107. Bueno/ bé/ yo si me tocara dinero bastante… (139.592 – bien)
108. Y esa época nosotros no la vamos a vivir porque ara ahora todo es… (218.289 – ahora)
109. Ese miedo ue produce el salir/ ara me meto/ tal/ esa fiesta me gusta (181.78 – ahora)
In the CSDCLP corpus, speakers use more discourse markers and with greater frequency than in PRESEEA. There are 19 instances of discourse markers used in the CSDCLP.

110. home/ si/ sí/ alguna habrá pero/ no sé a (lo) mejor son típicas de esas que a lo mejor vas toda acelerada/
111. y ara no entro
112. ara a nivel de: de loterías o de suertes de: de cosas/ pos sí que una vez tuve suerte
113. ¡y ya ves tú! ¡tú me ves ara lo que yo me pinto?
114. hicimos allí una picadilla/ algo de marisco muy baratito i avant//
115. no/ moro/ no/ una familia cargada i avant
116. y lo que es para uno pos se ve que está escrito que nos tiene que pasar/ i avant!
117. a vore, a los bomberos socorro socorro!
118. manillar alto con frenos de varilla/ pues la dejabas allí arrimá(d)a y no hay ninguna Harley ara ahí en/ en benicasim que la miraran más
119. ara ¡se dormía de bien encima!
120. ara porque les han quitado la camporita de los pubs
121. pos ara te diré ¡y no viste un alma!
122. ara/ si encontrara a lo mejor un trabajo de lo mío y sólo por las mañanas//
123. y: sí/ teníamos un susto bastante gordo// pero bueno cosas que pasan/ i avant
124. home/ a mí me parece que el curso escolar/ escolar/ debería durar lo que los psicólogos determinen que debe durar

The CSDCLP data shows more of a pattern than PRESEEA, as certain discourse markers evidence a more fixed usage. For example, the many instances of pos appear to function to separate thoughts. Conversely, a vore, ara, home, and això occur more often at the beginning of the sentence, although ara sometimes occurs medially as a discourse marker. Other markers, such as i avant, tend to occur in final position. The word pos is used extensively in the CSDCLP with 53 cases in the corpus.

125. pos me: me dejaron trabajar entonces ya ya terminé octavo y me puse directamente a trabajar.
126. pos el día que me separé
127. el destino o lo que sea/ pos si/
128. pos “¡el xiquet/ el xiquet!” ara no sé qué
129. pos hablar por hablar/ pero en fin/ no: en ese sentido/ yo creo que soy como mi padre un poco
130. y cada vez que hay: un maltrato a- como a esas mujeres y tal pos también me pone los pelos de punta/ que considero que el respeto a las personas es fundamental y si uno quiere que le respeten primero tiene que respetar
131. pos: sí que cuesta/ sí que cuesta/ hombre/ claro que cuesta
5.5.5.7 Polylanguaging into German, Portuguese, French and Italian

The Twitter data presents some clear instances of polylanguaging into other languages such as German, Portuguese, French or Italian. Some of the examples below are situational in nature, such as (132) and (133) below, where the use of German and Portuguese respectively is motivated by a prompt in the tweet. None of the examples of polylanguaging into other languages is found across users and each example is only found in one informant’s tweets. Therefore, it is difficult to argue that the user is proficient enough in any of the above languages to suggest something other than polylanguaging. Only (135) is syntactically complex, although only in comparison with the other examples. These are likely examples of polylanguaging as discussed by Jørgensen (2011). The users would not have to be proficient in any of the above languages to incorporate the following features into their tweets.

132. A los alemanes q conoci en el verano d 2010 y les conté cn alegría q habíamos ganado una euro y un mundial, ahora se lo merecen ellos! Danke [gracias] (F1A3)
133. Hannover hace oficial fichaje de Joao. "@user: Seja bem-vindo, (bienvenido) @user!: #H96 contrata João Pereira. " (M19V24)
134. Piove... [llueve] #valencia #Valenciagram #iger #igers #instapic #instagood #instadaily #igersValencia... https://instagram.com/p/0krycoxF3 / (F19V24)
135. Le plus grand. [Lo más grande] Mola Mogollón. Un tío grande. Allez allez mon ami. [Vaya vaya mi amigo] Felicidades @user (M34V40)

5.5.6 Established borrowings on Twitter

In looking at forms in the Twitter corpus which might be borrowings from Valencian into Valencian Spanish, issues of morphological adaptation was the starting point. Additionally if a word appeared either with more than one user or in more than one data source, then it was coded as a borrowing. At the same time, given the context and all that has been noted previously
regarding Twitter behavior, many of these might more accurately be described as currently instances of polylinguaging.

The following are likely nouns borrowed from Valencian into Spanish, describing the Falles celebration in Valencia.

136. Gente a la que le importa la crida: (M23G7)
137. Cuando es la mascletada esa (M23G7)
138. "@user: Entra el ninot d la falla San Josep de la Muntanya #ExpoNinot15 #fallas" ahí vamos @Calabazastudio (F16V23)

A few other items were determined to be borrowings due to their use by more than one user on Twitter as well as their being found in several corpora.

139. La banda d amiguetes d Torrente 5 me da más confianza que la banda del coleta. #PodemosNoGanara (F16V23)
140. xq gracias al Valencia te encontre, lo q una los colores q no los separe nunca @user #FF bonica Tú! http://youtu.be/qZWutsPAcwQ (F15V22)
141. "@user @user @user las más bonicas del pueblo me han dicho jajajaja" las mas beias del lugar tta (F13V18)
142. Seguirme en instagram bonicos : annapenelladuran5 (F11V16)

Bonico was used by more than one user and amiguete appear in more than one corpus.

The other single-word forms proposed as borrowings for the Twitter corpus are discourse markers, which in most instances here are Spanish discourse markers used in Valencian tweets rather than vice versa.

143. bueno me'n vaig a dormir que demà a les 07:00 estaré en peu! menut dia m'espera, 2 dies! :) #BonaNit (F1A3)
144. Air estaves per cullera potser? — Pues no jajaja, estava en Albal: ') http://ask.fm/a/b35qkk8o (M10V10)
145. Pues no puc dormiiir, musicaa pa' relajarse (M14V14)

Another form that appears to be a borrowing from Valencian into Spanish is mare.

146. mareee como lluevee (F8G8)
147. guuuuuapos mmare #myhyv (F8G8)
148. Mare meua tonight!!! (F13V18)
149. Nada,que no podemos ver un final tranquilos!! La puta mare! Vamos Valencia !!!! (F16V23)
There are a number of instances where a discourse marker from Spanish is inserted into a Valencian discourse, as seen in examples (143) and (144). Regarding bueno, researchers have noted its presence in other regional languages in Spain; Thomas (2008) notes its presence in Galician, while Vila i Moreno (1998) notes its presence in the Catalan of less competent speakers. The Valencian equivalent would be bo or bé (Gónzalez 1998; 2004). In Catalan, it appears frequently in informal conversations; however, it is not considered formal (Cuenca and Marín 2007: 907).

150. **bueno** me'n vaig a dormir que demà a les 07:00 estaré en peu! menut dia m'espera, 2 dies! :) #BonaNit
151. **Xe! Pues** se está posant interesant el partit! #BR (F2A4)

One interesting borrowing found on Twitter that does demonstrate morphological adaptation is the word hatear. This exists as a native word but the contexts in which it is used on Twitter indicate a borrowed meaning. The original word found in the RAE is ‘hatear’ meaning to ‘recoger la ropa y otros objetos de uso personal cuando se va a salir de viaje’ or ‘dar la hatería a los pastores’. However, considering the following examples, the intended meaning of hatear here is not the standard one.

152. **Ay dios, todos quieren q Alesso se la saque pero al minuto 1 lo van a hatear** #loveo (F1A3)
153. **No entiendo cuando hatean a Alesso** #Ultra2015 #UltraLive (F1A3)
154. **Aquí se hatea** la selección a partir de ahora. (M31V37)
155. **Hateadme** jajajaja asii nos vamos a comer mañana a los del Español ! Jajajajaja (F22V42)

Noun forms associated with the verb are also found in the corpus, albeit only masculine plural forms.

156. **No pueden haber haters** de este pedazo jugador, no pueden. (F19V29)
Given the context and the addition of the hashtag #loveo in (152), the intended meaning of this term is clearly to hate, especially in the context of modern culture and the use of the term haters in English-speaking culture to refer to discrimination against certain groups.

There are many English lexical items in the Twitter corpus where an analysis of the morphology of the noun indicates if it is a borrowing. In the following, the morphology reflects the degree of adaptation of the lexical borrowing. The use of masculine articles with gym suggests that the word has been appropriated into Spanish as a masculine, singular noun.

157. Cerrando la semana en el gym! (M5V1)
158. Rufete ficha hoy, pero aun está en el gym. (M17P1)

Other examples are the masculine nouns partner (as in tennis partner), piercing, picks, fair play adopted into Spanish. In each instance, a level of morpho-syntactic integration is suggested by the presence of Spanish possessive adjectives, masculine singular definite articles, or masculine plural demonstrative adjectives.

159. 6-4, 7-6 y ya en cuartos!! Esta tarde a por el dobles con mi partner @user #vamos (M5V2)
160. Empezamos la semana como acabamos.. Victoria 6-2, 6-4! Después del Match foto con mi partner de dobles @user (M5V2)
161. Se me ha roto la bola del piercing de la lengua. Qué raaaaro. (M8V5)
162. Por cierto, nadie le va a dar merito a @user, pero esas jugadas y esos picks (sobre todo primer game) vienen de su mano. Genio! (M28O1)
163. Bueno voy a echar un partido en la play a ver si pillo a un mandril y le reviento que ahí no expulsan a Otamendi, jajajajaja por ahora (M18V21)
164. Esta tarde voy a por @user acepto encargos :) que frikie soy (F15V22)

There are also instances where morphological adaptation is not formally evident but the word can be categorized as a borrowing. The following are examples of such cases.

165. Tranquila, tengo money y las hamburguesas no desaparecen. Puedes suicidarte tranquila ok, tk xao @user
166. Vaya touchdown (F20V30)
167. Puede hacerlo si aumenta los ingresos mediante patrocinadores o si hace ampliación de capital. Pero sí, hay fair play. @user (M34V40)
5.5.6.1 Nonce borrowings

5.5.6.2 Nouns

The corpus shows a number of Valencian nonce borrowings into Spanish. In the majority of these cases, a Valencian NP is inserted into Spanish discourse. There are a mixed variety of number and gender for the NPs. Poplack (2012, 2015) notes that nonce borrowings nouns are the most common type to occur due to the relative ease in which they can be switched and they are known as lone or single-word insertions.

168. De Viola lo que mas recuerdo es a "Violín"... A sus 46 años vuelve!!! Le ganó un coche a Paco Roig y se hizo la senyera en el pelo!!! :) (M19V24)
169. Estamos en semis amics [amigos]. (F19V29)
170. No os enamoréis amics. (F19V29)
171. Buenas noches, un beset [besito] @SevillaFC (F2030)
172. Yukismo en valenciano sería yúkisme que en inglés significa besame you kiss me osea que somos homosexuals (M23G7)
173. Querido @lazaroftherock está foto es mel [miel] de romero. Puxa, AMUNT (M34V40)

5.5.6.3 Verbs

Valencian verbs and verb phrases are included in the following examples. (174) is technically a verb phrase but it is behaving as a single unit due to the hashtag function. For this reason it was treated as a nonce borrowing.

174. #MeDanGanasDe pegarli una hostia a més de uno i no puc [pegarle una hostia además de uno y no puedo] (F2A4)

5.5.6.4 Adjectives

There are cases of code mixing with an adjective; for example, the following shows a Valencian adjective included in a Spanish tweet. Some adjectives that occur in the tweets of
various users and also across data sources, such as the word *bonico*, are most likely Valencian borrowings and not instances of lone or single-word insertions. (176) shows the insertion of an English adjective. This could be interpreted as an instance of polylanguaging.

175.  Estoy dando más vueltas en el 81 que en el bus *turístico*! Ya estamos en fallas!!!  
       (F16V23)
176.  O suicidio u otra cosa. Muy *heavy*. (M34V40)

### 5.5.6.5 Adverbs

The cases below show instances where Castilian adverbs are used in Valencian tweets. The issue of determining if these are cases of language mixing into Spanish by inserting these adverbs in Valencian text or if they are Spanish established borrowings integrated into Valencian is a difficult task. Here they are likely lone insertions due to the fact that they do not fulfill the requirement of frequency mentioned as a way of determining borrowings.

177.  No *mas* [més] fet ni puto cas!! xD @asmorsarot (F1A3)
178.  Com puc tindre tanta energia pel mati, *después* [després] de lo d'ahir??? (F2A4)

There are also cases of Valencian adverbs being used in Spanish tweets in (179) and vice versa, where a Spanish adverb is used in Valencian text in (180).

179.  *Dema* [mañana] a las 5 en pie, empieza lo bueno #rescue2014 #montpellier  
       (M14V14)
180.  *Después* [després] de la copa litrona a jugar a bàsquet, hui serà un dia mooolt llarg (M28O1)

### 5.6 Comparing lone vs. multiword insertions across data sources

This section analyzes the data with regard to the frequency of lone insertions versus the inclusion of multiword insertions. The data presented in the sections presenting the results from each data source reveal that there is a difference in preference with regard to lone insertions vs.
multiword insertions across data sources, for instance, the preference in Val.Es.Co for multiword insertions. The appearance of each type, lone insertions vs. multiword insertions, was quantified with regard to its use in each data source. Each instance of mixing was analyzed to determine if it involved lone and multiword insertions. Table 11 combines the different mixed language phenomena for a comparison between data sources. Single or lone insertions will typically refer to established and nonce borrowings, with the minimal use of intrasentential switching in the data included in the category of single insertions. Multiword insertions include the instances of CS and language mixing found in each data source.

Table 11. Single vs. multiword insertions

<table>
<thead>
<tr>
<th></th>
<th>Absolute frequency</th>
<th>Relative frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone insertions</td>
<td>292</td>
<td>76.24 %</td>
</tr>
<tr>
<td>Multiword insertions</td>
<td>91</td>
<td>23.75 %</td>
</tr>
<tr>
<td>PRESEEA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone insertions</td>
<td>96</td>
<td>80.67 %</td>
</tr>
<tr>
<td>Multiword insertions</td>
<td>23</td>
<td>19.32 %</td>
</tr>
<tr>
<td>CSDCLP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone insertions</td>
<td>124</td>
<td>76.54 %</td>
</tr>
<tr>
<td>Multiword insertions</td>
<td>38</td>
<td>23.45 %</td>
</tr>
<tr>
<td>Val.Es.Co</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone insertions</td>
<td>43</td>
<td>55.12 %</td>
</tr>
<tr>
<td>Multiword insertions</td>
<td>35</td>
<td>44.87 %</td>
</tr>
</tbody>
</table>

The results show a pattern across data sources with regard to the preference of lone insertions over that of multiword insertions. Twitter, CSDCLP, and PRESEEA have a preference for single word items, while the Val.Es.Co corpus shows less of a pronounced preference and shows even a slight preference for multiword insertions. The data differences between the PRESEEA and the CSDCLP are noteworthy as these two data sources are the most similar in that they both offer
data taken from sociolinguistic interviews. The conversational corpora also reveal a geographic
difference with regard to the level of Valencian spoken in the region. Castellón de la Plana
typically reports a higher self-reporting rate with regard to the use of Valencian, while the city of
Valencia reports an overall lower preference to use Valencian in public.

The preference for multiword insertions in Val.Es.Co is interesting. In seeking to explain
this preference, it must be noted that the one thing that separates these data sources, besides the
aforementioned geographic differences, is the data collection methods used in each corpus.
Val.Es.Co was gathered through secret recordings of conversations. Due to the nature of how
data was collected in the Val.Es.Co corpus, the results reflect a naturalistic, spontaneous
conversation, even more so than data deriving from a sociolinguistic interview. In Val.Es.Co,
there are no structured interview questions and the conversation follows a completely natural
pattern, due to the secret recordings, while the sociolinguistic interview structure used by the
PRESEEA and the CSDCLP, where an interviewer questions each ‘spot-lighted’ informant
individually. This interview structure leads to the interviewee providing short monologues in
response to the interviewer’s questions, thus creating a type of semi-spontaneous data collection
method. This difference between data collection protocols and their results suggests that there is
something that unites the PRESEEA and CSDCLP data with Twitter more than with Val.Es.Co.

It seems unlikely that speaker proficiency in Valencian would be the reason to explain the
differences in the data sources, although it is theoretically possible that informant selection could
explain the differences. If the issue of speaker competency is likely not driving the differences in
speech patterns here, then the explanation must lie with the nature of communication in each data
source. What is found is a continuum with regard to the degree of conversational turn taking, or
spontaneity, found in each data source with Val.Es.Co being more or less completely reflective
of a natural conversation and Twitter reflecting a completely non-conversational, non-spontaneous mode of communication, due to the issue of planning with regard to each post or tweet. The degree of turn taking appears to affect the pattern of mixing found in each source. In an informal conversation, speakers engage in turn taking as an integral part of communication. An interesting parallel between Twitter, PRESEEA and CSDCLP is their lack (Twitter) or reduced (PESEEA and CSDCLP) degree of turn taking. The established sociolinguistic interview protocol promotes little turn taking, less than that found in a typical conversation of rapid exchanges between sociolinguistic equals. On Twitter, a user broadcasts with no spontaneous interaction with others; hence, no real turn taking.

When turn taking is eliminated in communication, many of the social and pragmatic reasons for initiating CS are also eliminated as several of the motivations for CS are dependent on one or more peer interlocutors engaged in quick and spontaneous exchanges of information. This aspect is highlighted in the present study by the use of a different terminology to describe conversational code switching (CS) and written code switching (language mixing). This is due to the different processes that motivate an informant to include CS in conversation in comparison with the motivations for including language mixing in writing, due to the differences in the nature of the data as non-spontaneous in contrast to spontaneous production of CS and borrowing in bilingual discourse offline.

Grosjean's (1982) suggested list of motivations for CS offers an example of how an elimination or reduction in turn taking can undo many of the motivations for CS. Grosjean notes that language switching can happen when a person is speaking in a monologue-type of communication but that it is more common in a conversational (‘turn taking’) setting. The use of CS to indicate speaker involvement through a personalization of message is more salient in a
conversational setting as is the use of CS to exclude. It appears then that many of the identifiable motivations for phrasal CS are less applicable in CMC contexts or in those that involve less turn taking and spontaneity. Sebba (2012) suggests this, noting the tight correlation between CS and a conversational framework, as CS generally takes place in sequential fashion between utterances. With Twitter, an asynchronous mode of CMC, the nature of communication is non-sequential and marginally interactive. When a user tweets, it is a self-contained unit that does not need to be linked to another tweet. Grosjean's (1982) motivations for CS are presented below.

**Figure 5. Reasons for CS (Grosjean 1982)**

1. Fill a linguistic need for lexical item, set phrase, discourse marker, or sentence filler
2. Continue the last language used (triggering)
3. Quote someone
4. Specify addressee
5. Qualify message: amplify or emphasize ('topper' in argument)
6. Specify speaker involvement (personalize message)
7. Mark and emphasize group identity (solidarity)
8. Convey confidentiality, anger, annoyance
9. Exclude someone from conversation
10. Change role of speaker: raise status, add authority, show expertise

If turn taking is absent then (4), (6), and (9) are excluded as they deal with defining relationships in a conversational episode. (2) and (3) might still apply in theory on Twitter and other forms of CMC although the nature of Twitter as an asynchronous form of CMC makes (2) not particularly applicable and (3) is unlikely to occur due to the limitation of 140 characters per tweet. The 140 character limitation discourages users from using space quoting others and the presence of a built-in quoting function (the RT mechanism) eliminates the need.

Can Grosjean’s list (1982) be modified to take into consideration the differences between Twitter and offline conversation? The changes to (2) were discussed in the previous paragraph. (3) is proposed by Sebba (2012) and suggested by the results here (the importance of the use of language mixing online to create identities, which can be accomplished through the use of tags
and other formulaic expressions). (4) – (7) are maintained from the original list. The importance of (5) must be highlighted as users can express solidarity through the use of greetings and farewells and happy birthday wishes.

**Figure 6. Reasons for CS updated for Twitter**

1. Fill a linguistic need for lexical item, set phrase, discourse marker, or sentence filler
2. Alluding to another user or event but not directly quoting
3. “Identity creation in individual texts” (Sebba 2012: 11) via tags and other formulaic expressions
4. Qualify message: amplify or emphasize (‘topper’ in argument)
5. Mark and emphasize group identity (solidarity)
6. Convey confidentiality, anger, annoyance
7. Change role of speaker: raise status, add authority, show expertise

The revision to Grosjean’s (1982) list above shows that in a non-conversational setting, language mixing loses conversational properties. While in a written setting language mixing still retains some aspects of communicative conversational CS, the function of language mixing in this context is more accurately described as performative either by performing identity or adding an expressive element to communication. The issue of polylanguaging must be contextualized with regard to this list. It is likely that the function of polylanguaging mirrors that of language mixing although the issue separating these phenomena is not one of function but of proficiency and complexity.

In synchronous modes of CMC, exchanges between users occur more rapidly and with a higher rate of turn taking, which more closely resemble oral conversation. In asynchronous modes of CMC such as Twitter, the period of time between individual contributions is longer and is less similar to much of naturalistic conversation. Paolillo (2011) hypothesizes that synchronous modes of CMC will have more conversational language mixing. Both Paolillo (2011) and Lee (2007) confirm that CS is more likely to be found in synchronous CMC modes. However, Hinrichs (2006) notes that creative language mixing can occur in asynchronous modes.
as well. Androutsopoulos (2011) theorizes that each CMC mode offers users different opportunities for bilingual discourse. Previous research and the suggestions made by this study suggest that language mixing and polylanguaging in a non-conversational or asynchronous setting favors more a performative function than a communicative or conversational one.

Other differences should be noted regarding communication and mixing patterns online and offline. The issue of planning is important in asynchronous modes of CMC where there is considerable time between interactions, which affects the type of language mixing produced. The lack of spontaneity specifically on Twitter suggests that CS online (‘language mixing’) should be thought of as a different process from CS in offline conversation.

Some CMC researchers note that users seem to avoid or prefer certain types of switching, “The metalinguistic awareness involved in planning and editing posts may inhibit the spontaneous, unconscious process of code-mixing, and the public character (and ethnolinguistic heterogeneity) of these forums is at odds with the situational conditions that favour the occurrence of code-mixing” (Androutsopoulos 2011: 17). However, the opposite has also been found, i.e. that planning allowed users the ability to be more creative than they would be in natural speech, which would promote more instances of polylanguaging (Jørgensen 2008). The level of polylanguaging however found in Jørgensen’s (2008) data does not occur at the same level as that found in the Twitter data here.

The effects of planning versus spontaneity however is not a black or white issue and Androutsopoulos’s (2011) assertion that planning limits language mixing does not necessarily mean that the lack of spontaneity limits the extent of language mixing but only the type of language mixing found. The lack of planning might explain the proliferance of tag switching on Twitter due to the fact that these tags are not complicated and do not require much proficiency or
creativity. Thus affirming that perhaps in a medium, such as Twitter, where there is typically more planning associated with posting than in other more synchronous mediums, such as instant messaging, there will be less cases of spontaneous CS in the form of intrasentential switching, there could be an increase in emblematic switching in tags. The data from Twitter suggests that Valencian Twitter users do rely on emblematic switches for the various reasons proposed above.

Regarding the issue of lone vs. multiword insertions and the disparities between Twitter, CSDCLP and PRESEEA additional possible explanations are the inclusion of instances of polylanguaging in Twitter data which would skew results towards lone insertions and the demographics on Twitter users which skews towards a younger informant group who arguably have a less fixed linguistic repertoire and are perhaps more open to linguistic ‘play’.

5.7 Language use per individual informants

The following table shows how many informants from each data set incorporate CS general language mixing and polylanguaging into their speech. Although the difference between data sets in terms of informant selection, data collection protocol and context of communication problematizes any simple comparison of informant totals, it is still worth noting the differing totals and what they say about the effect of the different mediums and conditions of communication vis-à-vis the studied contact phenomena.

**Table 12.** Number of informants with language contact phenomena per medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Number of informants who mix</th>
<th>Total number of informants</th>
<th>Percentage of informants who mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>47</td>
<td>60</td>
<td>78.33 %</td>
</tr>
<tr>
<td>PRESEEA</td>
<td>29</td>
<td>72</td>
<td>40.27 %</td>
</tr>
<tr>
<td>CSDCLP</td>
<td>42</td>
<td>72</td>
<td>58.33 %</td>
</tr>
<tr>
<td>Val.Es.Co</td>
<td>29</td>
<td>196</td>
<td>14.79 %</td>
</tr>
</tbody>
</table>
As the table shows, the distribution of mixing among informants is different within the different data sets. With Twitter, there is an overall higher rate of the studied contact phenomena. This contrasts most significantly with the Val.Es.Co data, where the percentage of informants who evidence a remarkably low mixing pattern, particularly given everything mentioned earlier regarding their data collection protocol being friendlier towards such phenomena occurring. In this case, it is important to note the higher rate of multiword insertions in the Val.Es.Co data. Instances of a higher rate of nonce borrowings in the PRESEEA and the CSDCLP might explain a possible inflation of their final totals, as well as differences between those two corpora and the Val.Es.Co with regard to informant selection.

An additional issue to consider with regard to the presence of language mixing is its distribution among informants. Are there certain users/informants who show much higher instances of CS than the average and subsequently skew the results? A series of graphs showing the number of instances of CS for each informant is shown below.

**Figure 7. Distribution of CS/language mixing among Twitter users**
In this graph specifically of Twitter informants, it can be seen that one user in particular shows a quite high number of instances of language mixing. However, a few things might contribute to this. The main thing to consider is that each Twitter informant inputs a drastically different number of tweets. Some are casual users who post only occasionally, some are semi-regular and others are hyperactive users who post several tweets per day. This latter group is usually promoting some kind of self-image, which could explain why some users have higher numbers of instances of mixing or polylanguaging.

**Figure 8. Distribution among informants in the PRESEEA corpus**

In the PRESEEA corpus, the differences among informants with regard to the rate of frequency specifically of CS is much less pronounced than that of language mixing and polylanguaging on Twitter with the speaker with the most cases of CS not drastically exceeding that found for several other speakers.

**Figure 9. Distribution of CS among informants in the CSDCLP**
The graph representing the CSDCLP informants resembles Twitter more than it does PRESEEA in that there is one informant with a higher number of occurrences of CS.

**Figure 10.** *Distribution of CS among informants in Val.Es.Co*

Do informants with disproportionately higher frequencies of CS demonstrate any identifiable characteristics that distinguish them from other informants?
In terms of the Twitter informants, users with more than 30 occurrences of language mixing in their profiles evidence patterns found for other user profiles such as tag switches and other intrasentential switches. One thing that does differentiate these users is a more prominent use of Valencian and the use of Spanish borrowed terms or switching into Valencian. The following examples will show some instances from the data:

181. **bueno** me'n vaig a dormir que demà a les 07:00 estaré en peu! menut dia m'espera, 2 dies! :) #BonaNit (F1A3)
182. Te collons el @HospitalRibera, criden x telefon x a cambiarme la cita a un dia antes xo no me la cambien x a un altre dia, Putos! (F1A3)
183. No **mas** fet ni puto cas!! xD @asmorsarot (F1A3)
184. Com puc tindre tanta energia pel matí, **después** de lo d'ahir??? (F2A4)
185. On hi ha una discoteca oberta? **Què voooooooooooom** (F2A4)
186. Jo que volia anar-me'n contenta a dormir... i al final encara ha empatat Portugal... **2-2** #USAvsPOR #shit (F2A4)
187. “@user @user che com te cuides me caguen deula jajaja” **ta to pagao!! break en el curro** (F15V22)
188. @user @user home si son iguals, es el meu cosí perdut **lo menos** (F22V42)

However, these same types of switches are found in other profiles as well:

189. **Pa que después** digen de les xiques... asi está una parlant de fútbol amb els xics :P (M3G3)
190. Booooondia al costat del **goordito** #RETRETA (M10V10)
191. Cuando tengas 90 años, ¿qué es lo que más te importará? — Anarmen de festa jejejeje **http://ask.fm/a/b1lq3ej** (M10V10)
192. "@user: Te necesito como el aire para respirar.26 **pic.twitter.com/TLp59mMgD" no te puc voler mes carinyo** (M10V10)
193. **Después** de la copa litrona a jugar a bàsquet, hui serà un dia mooolt llarg (M28O1)

The Twitter data shows that there is a slight increase in the use of Valencian as the matrix language in cases of language mixing in the profiles with a larger number of occurrences of it overall. The atypical speakers in the other corpora do not demonstrate any discernible pattern that would differentiate them from other informants.
5.8 English in the conversational corpora and Twitter

One additional issue to be addressed is the presence of English on Twitter. Nearly 35% of the cases of mixing on Twitter contain English insertions. The issue of borrowings on Twitter was discussed, particularly in relation to the incorporation of English. Cases of English use in hashtags are interpreted as polylanguaging in this study for reasons mentioned earlier. Additionally, other lone insertions are used that are likely established borrowings due to their morphological adaptation into Castilian and their frequency across data sources or informants. The presence of English on Twitter, whether as polylanguaging or with borrowings is noteworthy.

Table 13. Most frequently used borrowed terms in the data sources

<table>
<thead>
<tr>
<th>Term</th>
<th>Twitter</th>
<th>CSDCLP</th>
<th>Val.Es.Co</th>
<th>PRESEEA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hobby</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Crack</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Hatear</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Relax</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Gym</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Partner</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Haters</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Camping</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Show</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Marketing</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Chat</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Other single word insertions</td>
<td>14</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>23</td>
</tr>
</tbody>
</table>

Included in the totals are variants such as conjugated forms, singular or plural forms and alternate spellings. Most of the forms above are only found in the Twitter corpus. This might reflect the nature of the Internet as a contact zone where new words might be more easily introduced and spread. Additionally, the nature of Twitter as an egocentric CMC medium
encourages creativity as well as self-consciousness from users. The –ing morpheme, although found in several data sources, does not appear with any words shared among them. The morpheme itself does not appear to be borrowed but rather words with the morpheme are, occupying a semantic field related to a physical activity. Examples include camping, running, tumbling, zapping and footing.

5.9 Linguistic economy on Twitter

Linguistic economy is potentially a determining factor in language use on Twitter, given the imposed limit of 140 characters per tweet. Naturally this distinguishes language use on Twitter from conversational corpora. Economy as a motivating factor would correspond to a user self-consciously selecting words in an effort to reduce the number of characters.

In order to test the effect of linguistic economy, instances of lone insertions from the Twitter corpus were tested comparing elements switched in order to determine if an equivalent in the base language of the tweet contains more or fewer characters. For example, bona nit would be compared with buenas noches. As Bona nit contains 7 letters it will require fewer of the 140 characters.

“supongo que vendrán mejores momentos, bona nit”

In the Twitter corpus 115 of the 383 instances of lone insertions were examined. Some tags were analyzed, such as the example above. Duplicates of the same tags were not doubled-counted. In the section above on tag switches

Shin (2010) analyzed the effect of linguistic economy on lexical borrowing in instances where bilingual speakers use elements where “borrowing is not due to the need to fill a conceptual gap” (46). Her argument is that in cases where speakers are not borrowing to fill a
gap (lexical need) “one factor that helps determine which words from the donor language will replace equivalent terms in the recipient language is efficiency, as measured by word-length” (46 – italics mine). Her research shows that English lexical insertions into New York Spanish are significantly shorter than their Spanish counterparts, supporting the assertion that “efficiency, as manifested by word-shortening, is a significant factor that influences lexical borrowing behavior” (56).

This study proposes that efficiency could also affect lone insertions on Twitter, as well as with some tags, which are ‘multiword’, yet emblematic in nature. Tags and other instances of polylanguaging many times involve lon insertions. This study examined to see if Shin’s findings might extend also to two typologically similar languages, Spanish and Valencian. Shin (2010) argues that overall, Spanish words are longer than English words. In order to determine if Valencian words are generally shorter than Spanish words, Spanish-Valencian word pairs were selected from the Twitter corpus, which did not appear in instances of language mixing. A sample of 83 word pairs were analyzed applying Shin’s (2010) methodology as a model. When the number of cases of equal grapheme length is excluded from the sample, we see that Valencian words are more likely to be shorter in terms of syllables than Castilian words (82.14% to 17.85%) and orthographically shorter (64% to 36%). However, it seems that, overall, Valencian and Spanish words are roughly equivalent in length in terms of the number of characters.

Table 14. Valencian words in comparison with Castilian

<table>
<thead>
<tr>
<th></th>
<th>Number of cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valencian word is longer</td>
<td>18</td>
<td>21.68 %</td>
</tr>
<tr>
<td>Valencian word is equal</td>
<td>33</td>
<td>39.75 %</td>
</tr>
<tr>
<td>length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valencian word is shorter</td>
<td>32</td>
<td>38.55 %</td>
</tr>
</tbody>
</table>
Looking at the Twitter data where switching does occur into Valencian, the following table summarizes the potential importance of economy as a motivating factor encouraging switching into Valencian.

**Table 15. Orthographic economy in lone insertions on Twitter**

<table>
<thead>
<tr>
<th>Data source</th>
<th>Number of occurrences</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>More characters in lone insertion than equivalent</td>
<td>21 cases</td>
<td>18.26 % of the cases</td>
</tr>
<tr>
<td>Equal number of characters in lone insertion than equivalent</td>
<td>26 cases</td>
<td>22.6 % of the cases</td>
</tr>
<tr>
<td>Fewer characters in lone insertion than equivalent</td>
<td>70 cases</td>
<td>60.86 % of the cases</td>
</tr>
</tbody>
</table>

The table suggests that, *at least in terms of number of characters*, users may often be manipulating language choice with linguistic economy in mind.

### 5.10 Language use in conversational corpora and on Twitter

As noted in the above sections describing the results of this study, one general result that characterizes all of the data sources is the infrequency with which Valencian Spanish speakers use Valencian in their discourse. In the conversational corpora, there were only 359 tokens of any mixed language phenomena, including CS, nonce borrowings, and established loanwords. The corpus of spoken Valencian Spanish compiled to analyze in this study included over one million words, 187 sociolinguistic interviews and 340 speakers of oral data in comparison with the written data from Twitter comprising data from 60 Twitter users from Valencia with 106,719 words and 8,432 tweets analyzed. This represents a tiny fragment of the total number of words actually spoken in this corpus. In the Twitter data, a relative frequency can more easily be established comparing the number of tweets with any type of insertion from Valencian to see that...
around 3% of the tweets are composed entirely in Valencian and only 1.6% of the tweets contain Valencian insertions (3% with all insertions from English and Valencian), with a majority of those cases being emblematic uses of Valencian on Twitter. This data suggests that Valencian is used in a similar way to other regional languages, such as Irish and Galician, as studied by O’Rourke (2005). She notes that there is a connection between language and identity but that this link is “be little more than a superficial marker of identity, and positive support for the language on this level need never move beyond its symbolic role” (277). The pattern of use of Valencian in the data suggests a largely symbolic value assigned to Valencian, which does not equate actual use by speakers, except in occasional emblematic uses, particularly on Twitter, an “associated language” (Eastman 1984), which has symbolic power to signify value distinct from actual use. O’Rourke (2005) found similar patterns with regard to Irish and Galician and others found a similar pattern with Welsh (Keyes 1981). This pattern suggests a similar pattern of adopting an historical, regional language as an associated language. The case of Welsh and Galician share with Valencian the lack of a strong political independence movement as found in Catalonia.

5.11 Conclusion

This study analyzed the presence and use of lexical borrowings and code switching in Valencian Spanish in two diverse communicative contexts, as well as the phenomenon of polylanguaging in English hashtags.

The overall preference in both the offline and online data was for Spanish as the preferred code (Giacalone-Ramat 1995), which corresponds to the recent language use surveys from the region. The use of English, arguably the ‘international language’ of social media, is
highly prominent on Twitter, presumably an effect of polylanguaging, in contrast to the conversational corpora where only English established borrowings are found; the highest frequency of English borrowings was found in CSDCLP. An additional observation is that in the conversational corpora there are more cases of code switching to mark reported speech.

Each conversational corpus had its own unique patterns with regard to the appearance of code switching and lexical borrowings. The PRESEEA corpus showed only a few instances of Valencian intrasentential nouns and in terms of CS, the majority of instances were of clauses. The CSDCLP also prominently shows Valencian established and nonce borrowings in single-noun insertions and with discourse markers. In terms of CS, the CSDCLP shows more instances of intersentential switching and more instances of CS than what is found in the PRESEEA. In the Val.Es.Co data there are a higher number of cases of CS with verbs than in other corpora. It also has notably more instances of multiword insertions overall in comparison with other data sources.

Regarding the Twitter data, the most frequent type of switch found are tag switches, which appear in a highly emblematic fashion. These include exhortatives, greetings and farewells, happy birthday wishes, and a variety of other types of tags. After tag switches, in terms of frequency, users apply intrasentential switches and then intersentential switches. Individual nouns and noun phrases are the most switched syntactic constituents, followed by discourse markers, adverbs, and verb phrases.

Regarding the use of lone vs. multiword insertions, data sources differed from one another in noteworthy ways. Twitter, CSDCLP, and PRESEEA shared a strong preference for lone insertions. The Val.Es.Co data revealed a preference for multiword insertions, a difference likely due to the data collection methodology used.
On Twitter, English is included in 35% of the tweets, which is at a high rate of frequency although at a lower frequency than what appears for Valencian. It was suggested that the use of English on Twitter and in other data sources was likely not a case of CS/language mixing, but either of established lexical borrowing or polylanguaging. On the other hand, the use of Valencian as the ‘embedded’ language constituted either instances of borrowing or CS in the conversational corpora and Twitter.

Linguistic economy was also considered as a motivation for language mixing on Twitter by comparing between languages the number of graphemes needed for equivalent expressions or words. Findings indicated that users selected the shortest lexical item available from their personal linguistic resources in 60.86% of the cases analyzed. The consideration of grapheme economy was considered due to the technological limitation on Twitter of 140 characters per tweet.
CHAPTER SIX
CONCLUSION

6.1 Introduction

This final chapter presents a brief summary of the study’s findings and answers to the research questions laid out in the first chapter.

The relevance of this research will be summarized with regard to furthering our understanding of how speakers differently apply contact phenomena (linguistic borrowings, CS, polylanguaging) associated primarily with either offline or online acts of communication. This study additionally analyzed the interplay between a regional and national language, Valencian and Spanish, and how each is used online and offline. The differing use of each language reveals how speakers relate to one another in very different ways according to medium and communicative context. This chapter will also outline some of the limitations of this study and present some suggestions, based on this research, for further related avenues of research.

6.2 Summary of Findings

The research questions announced at the start of the study were the following:

1. In Valencian Spanish, what patterns exist of CS and lexical borrowing from Valencian in terms of constituent class of switched elements, of intrasentential, intersentential and tag switching, and of phrasal vs. single-word switching?
2. In what ways do the three published corpora of oral interviews offer differing results regarding the contact phenomena mentioned above?
3. How do the results from the three published corpora of oral interviews differ from those found in the corpus of Twitter data?
4. What factors explain any noted differences between the four corpora (online and offline)?
5. To what extent are concepts associated with an analysis of contact phenomena of spoken speech applicable to an analysis of similar phenomena found in Twitter data?

This study analyzed the presence and use of lexical borrowings, both established and nonce, and code switching in Valencian Spanish in two diverse communicative contexts (oral conversation and writing mediated by technology) as well as considered the phenomenon of polylinguaging as it relates to Twitter, especially in English hashtags. Essential differences were presented between communication online and offline in terms of how speakers use their linguistic resources and the different motivations, which determine code selection. A comparison between online and offline speech adds to the developing field of CMC research by comparing data taken from Twitter with that taken from conversational corpora.

In the geographic/cultural context which produced the data here, the overall preference in both the offline and online data was for Spanish as the preferred code, which is to be expected based on the linguistic situation in Valencia. On Twitter one sees more variation in language use and preferred code, with more use of Valencian in tweets with Spanish discourse markers or with other Spanish insertions. The use of English, arguably the ‘international language’ of social media, is prominent on Twitter as an effect of polylinguaging. An additional observation is that in the conversational corpora there are more cases of code switching to mark reported speech, as Twitter both has a retweet function that serves that purpose and the 140 character limit per tweet discourages its ego-centric users from wasting space reporting the speech of others.

Each conversational corpus had its own unique patterns vis-à-vis the appearance of code switching and lexical borrowings. The PRESEEA corpus showed only a few instances of Valencian intrasentential nouns and in terms of CS, the majority of instances were of intersentential type. The CSDCLP also prominently shows Valencian established and nonce
borrowings and with discourse markers. In terms of CS, the CSDCLP shows more instances of intersentential switching and more instances of CS than what is found in the PRESEEA. In the Val.Es.Co data speakers CS with at a much higher frequency than what appears in the other data sources. It also has notably more instances of multiword insertions, perhaps due to its particular data collection technique which promoted a level of conversational turn-taking.

Regarding the Twitter data, the most frequent type of switch found is tag switch, which appear in a highly emblematic fashion. These include exhortatives, greetings and farewells, happy birthday wishes, and a variety of other types of tags. After tag switches, in terms of frequency, users apply intrasentential switches and then intersentential switches. Individual nouns and noun phrases are the most inserted syntactic constituents, followed by discourse markers, adverbs, and verb phrases. One notable finding is the high frequency of tag switches in the Twitter data. They do not occur as frequently in the conversational corpora as on Twitter where they have at least two identifiable functions. The first relates to identity creation through language mixing (Sebba 2012), and the second to being a way of establishing solidarity with other users, particularly regarding the language used in greetings and farewells. The published corpora do not exhibit the same degree of tag switching.

The Twitter data opens up questions about how discourse is organized differently offline and online, which affects categorizing and comparing instances of intersentential and intrasentential switching. The use of hashtags and how they fit into online discourse is complicated. As with reported speech, in some, albeit rare, instances hashtag material appears syntactically and semantically integrated into the rest of the tweet.

Beyond offering a general characterization of the data collected from published corpora and from Twitter, the study analyzed factors that influence the appearance of linguistic
borrowings and code switching. The degree of complexity and issues at stake associated with the different types of switches was discussed. The presence or absence of conversational turn taking was suggested as a possible explanation for differences between data set findings. Regarding the use of single vs. multiword insertions, data sources differed from one another in noteworthy ways. Twitter, PRESEEA and the CSDCLP shared a strong preference for single-word insertions. The Val.Es.Co data revealed a preference for multiword insertions, a difference likely due to the data collection methodology used. One thing that the first three data sources have in common (Twitter, CSDCLP, and PRESEEA) is a structural lack of conversational turn taking. While this is obvious in the Twitter data, it is also true for the most part within structured sociolinguistic interviews, where the interviewer purposefully does not participate in the conversation but allows the informant to produce usually brief monologue-like responses to a given question/topic. The Twitter data reflects essentially a parallel context, where a user continuously presents very short ‘monologues’ and does not engage in turn taking with an identifiable interlocutor. The lack of turn taking in the first three sources is in contrast to the Val.Es.Co which employed a less structured (secretly recorded) data collection technique, which captured data truer to that found in natural conversation.

On Twitter, English is included in 35% of the tweets, which contain polylinguaging, which is at a high rate of frequency although at a lower frequency than what appears for Valencian. It was suggested that the use of English on Twitter and in other data sources was likely not a case of CS/language mixing, but either of lexical borrowing or polylinguaging. On the other hand, Valencian insertions on Twitter were coded as either cases of nonce or established borrowings or language mixing, if occurring as a multiword insertion.
Linguistic economy was also considered as a motivation for language mixing on Twitter by comparing between languages the number of graphemes needed for equivalent expressions or words. Findings indicated that users selected the shortest lexical item available from their personal linguistic resources in 60.86 % of the cases analyzed. The consideration of grapheme economy was considered due to the technological limitation on Twitter of 140 characters per tweet.

Examining patterns of both CS, lexical borrowing and poly languaging in both offline and online communication in Valencian Spanish highlights differences between these two modes of communication in particular as they relate to the underlying pragmatic goals that a speaker has in each situation.

6.3 Limitations of this study

There are limitations with using published conversational corpora as well as with using online data. Some derive from using data from other researchers’ corpora, as one is limited by the data that is available in each corpus, by the number of speakers interviewed, by the informant selection criteria, and by possibly differing data collection techniques. Although all of the corpora used here, with the exception of the Val.Es.Co, include some sociolinguistic information about each informant, each corpus used a different system and sometimes information included in one corpus was not included in another. This problematized a fuller investigation of extralinguistic factors and a more exacting comparison between sources regarding the effect of communicative context.

A second limitation of study was the reliance on probably dissimilar populations among informants, as the group of informants providing data for the conversational corpora is arguably
a rather different group of people than the informants who produced the data for the Twitter corpus. A study that could analyze the speech patterns of one population online compared to a similar population offline would provide a clearer understanding of the particulars of language selection of that group in their online and offline outputs as well as, more broadly, how language differs based on the context of communication.

Distinguishing between instances of CS and borrowing was a challenging issue with regard to classifying mixed language use on Twitter. This study relied on two criteria to determine if the inclusion of a single-word item from another language was a borrowing or an instance of CS. The first criterion was the presence of morphological adaptation via the inclusion of determiners or other morphology in the matrix language. The second criterion was the rate of frequency with which a particular term was used in the data sources. If more than one informant used a word and especially if it was used in more than one data source then the item was classified as a borrowing.

Another issue regarding categorization with the Twitter data was the determination of clausal boundaries in terms of a tweet, given the presence of hashtags and other features unique to online discourse. The position adopted was to consider messages, posts, and hashtags as “base-level units” (Androtsoupolous 2013: 237). This is particularly consequential in establishing the language mixing phenomena on Twitter.

6.4 Directions in future research

Issues of CS and borrowing have been considered in other studies on CS from a variety of population samples. This study adds to the field of CMC by considering the interaction of a regional language with a majority language and the interaction of both of those with a universal
language. The comparison of CMC data to non-CMC data in the same study is underrepresented in current CMC research. This comparison of CMC and non-CMC data can be expanded to consider other aspects of CS and borrowing. For instance, the issue mentioned above of using the same population for a point of comparison could allow a researcher to understand more fully how users vary in their use of language mixing and borrowing online and offline.

Another focus for future research would be the comparison of different groups of bilingual speakers to understand if there are general patterns with regard to the use of minority languages on social media. For instance, studies on other regional Spanish languages, such as Galician and Euskera could be conducted to understand if other populations in Spain share the same patterns vis-à-vis their inclusion of a regional language and English in conjunction with Spanish. Other populations could be studied from Latin America, such as Quechua-Spanish bilinguals or Guaraní-Spanish bilinguals. In this sense, the interaction between Castilian and indigenous languages could be analyzed via social media sites.


Auer, Peter. 2000. Why should we and how can we determine the “base language” of a bilingual conversation?” *Estudios de sociolingüística*, 1(1): 129-144.


Blas Arroyo, José Luis. 1999. Las actitudes hacia el bilingüismo en las comunidades de habla hispánicas. *Lebende Sprachen* 3. 120-129.


Calsamiglia, Helena and Empar Tuson. 1980. Ús i alternança de llengües en grups de joves d'un barri de Barcelona: Sant Andreu de Palomar. *Treballs de Sociolingüística catalana* 3. 11-82.


Grefenstette, Gregory and Pasi Tapanainen. 1994. What is a word, what is a sentence? Problems of tokenization. 3rd Conference on Computational Lexicography and Text Research.


Moll, Frances de Borja. 1961. El castellano en Majorca. Homenaje ofrecido a Damoso Alonso; por sus amigos y discípulos con ocasió de su 60a aniversario, 2(1), 469-475.


Wikström, Peter. 2014. #srynotfunny: Communicative functions of Hashtags on Twitter. SKY Journal of Linguistics 27. 127-152.


APPENDIX A: TAGS USED ON TWITTER

1. Buenisimas noches a todos, me voy con el mejor recuerdo, el set de Alesso hoy en el ultra miami #bonanit
2. hablando y hablando se me hacen las 3, mañana estaré fuera de cobertura todo el día... #bonanit
3. A dormir que mañana no tengo nada que hacer, solo disfrutar del relax y del solecitoouoo! #BonaNit
4. Buenas noches user! que hacia días que no te lo decía... descansa mucho y cierra los ojitos #BonaNit @user
5. 13 horitas para ver el mar y disfrutar del sol! ^_^ #BonaNit
6. A descansar que hay sueño y mañana peluquería :) #BonaNit
7. Siempre duermo con ella y por siempre #OrgullosoDeSerValencianista #AmuntValencia
8. Llegamos a casa, bona nit
9. Como paso de ver a Guetta y Skrillex... una que va ya a la cama #BonaNit
10. y con buena música de fondo, me voy a dormir! #BonaNit
11. No ha estado mal la tarde por Valencia... #BonaNit
12. Buaah ahora va y no me retransmiten el set de Tiësto ... :'( a dormir #BonaNit
13. AdioosMessi, como tu falta... bye bye
14. Quedan menos de 10 minutos para q empiece Hardwell en el #UltraEurope ... pero estoy muerta! mañana ya veremos su sesión :) #BonaNit
15. Me encanta ver el cielo por la noche #relax #GoodNight :
16. Menos mal que mañana ya estaré en la playa... #querelax #BonaNit :
17. ayer me quede sin bateria y no pude contestar... espero q hayas dormido bien #BonDia @user
18. Llevo cargando el móvil dos dias seguidos... #MeCabreo #BonaNit
19. Después del peliculón de Wall Street a dormir y soñar con mi amore Shia ❤ #BonaNit
20. A por el lunes con energíaaaaa!!! Bon diaaaaa
21. Adiós al weekend y hola a una gran y esperada semana. :) #BonaNit
22. Lunes intenso, diferente y lleno de emociones. Bona nit
23. Tengo sitio en el parking con sombra!!!!! #goodmorning
24. Keep calm And catch the mondays #BuenosDias #APorElLunes ya falta un día menos @user@user @user
25. Buenos días!! Bon dia!! Me encanta vivir en el Mediterraneo. BENICARLO!!
26. "@Mister_CeS: Buenas noches a toda mi gente con SUAJ." @user good night baby
27. Hacer la siesta boonanit guap@s ;)
28. Viive el presente y no mires el futuro.. #Boonanit
29. En fin bona niiti
30. Una que se ba a dormir <3//bona nit guapuritas!
31. Bona nit. #Desconexión
32. Después de ver Vive Cantando a dormir que mañana hay que madrugar para hacer la matrícula, bona niiiiiiit
33. Bye bye pelo largo!
34. Cuando un sentimiento habla por si solo ya no hay nada que hacer, no habrá nada capaz de pararlo #Bonavesperada
35. **bona nit**, mañana primer día de clase
36. supongo que vendrán mejores momentos, **bona nit**
37. **Bona nit .y Amunt!!!!** Próximo objetivo ,nuestros amigos de Elche.
38. Jornada resuelta, tres puntos importantes fuera de casa. No jugamos bien pero ganamos, vamos progresando. El próximo en Mestalla. #amunt
39. Lista y preparada !!!!! Vamos vamos ganas de Valencia !!!! Hoy 1-2 y a la terreta ! #Amunt
40. Buuuuenos días mi gente ! Ya mareando x aquí DISFRUTAR DEL SABADETE! #amunt Valencia club de futttbol lolololololo a las 18h tenemos cita
41. Que paséis un viernes LLENO DE SONRISAS! #amunt
42. Buenas Noches Valencianistas #Amunt
43. Las noches volverán llenas de luz, de magía y de color a Mestalla #JuntsTornem
44. día de poco fútbol pero victoria importante por los 3 puntos, compañía 10 con @user y @user #LateNayra #amunt
45. @user tres puntos que vienen de perlas.amunt.
46. “@user: @user Muchas Felicidades!! Disfruta de tu día, con los tuyos jajajaja AMUNT!” muchas gracias amunt !*
47. “@user: @user felicidades valencianismo. Q disfrutes de tu día.” Muchas gracias amunt SIEMPRE !*
48. “@user: Feliz cumpleaños @user se van notando ya las canas? ” gracias Crack! amunt !!!*
49. “@user: @user Muchas felicidades Maria! Disfruta de tu día y Amunt Valencia!! :)” muchas gracias AMUNT !*
50. “@user: @user Bondia guapa ” buenos días File!!
51. **Bonanit** El viernes #QuieroSerTercero
52. “@user: @user //Claaaaaaaro que lo se, y ellos también!!!/ **Bona nit guapa/AMUNT” ** Bonanit Number one
53. Buenas noches Valencianistas. Ser del valencia es el mejor regalo que me ha hecho mi familia. #BonaNit
54. “ @I_rechola : @merimachan **bona nit guapa**, felices sueños, mañana mas,” **Bonanit** que descanses
55. “@user: @user Buenas noches Maria mañana más y mejor!!” **Bonanit** Carlos!!
56. **Bon día** lluvioso :) un día menos xa ver a nuestro Valencia
57. Las noches volverán llenas de luz, de magia y de color a Mestalla #JuntsTornem
58. “ @file_013: @merimachan Bondia guapa ” buenos días File!!
59. “ ( @soundtecdigital: @merimachan Bdia Maria/FELICIDADESFFFFSSSSS :))//QTNGASGRANDIA !!! ” GRACIAS JOSE!! sisi gran día!!
60. Que importante ha sido el gol de Negredo hoy. Jornada casi perfecta. Recortando puntos. #ObjetivoChampions #JuntsTornem
61. **BonaNit //Al despertar me acuerdo de ti,por el día siempre te veo en algún lugar y mi último pensamiento es volver a verte jugar #juntsTornem
62. Moltes felicitats pels 28 @user !! passa un bon dia #HappyBday :
63. Hoy cumple 29 años @MichaelPhelps Felicidades!! #HappyBirthdayMichaelPhelps
64. 29 años para @MichaelPhelps Felicidades ! #HappyBirthdayMichaelPhelps
65. **Bona nit** :) sin mas, estoy reventada!

66. “@user: [user] vamos que nos vamos!! **Bonanit** carinyet!!!” Q ganitassssss!!!!!

67. **Bona nit** :) deseadando que sea domingo y descontando los días xa ir a Elche. #23dias #amunt

68. #GoodMorning #buenosDias [https://instagram.com/p/zeqJAvK2tK/](https://instagram.com/p/zeqJAvK2tK/)

69. **Bona nit lluneta.”** Bonanit

71. “@user: [user] #sed*, María. Te corrijo porque te quiero. **Bona nit lluneta.”** Bonanit

72. “@user: [user] Buenas noches amigas.” guapas que descanséis bonaniit

73. **Bon día!!** 2 días para volver a Mestalla!!

74. Toca ir a dormir, que mañana hay que levantarse prontito gracias a Tebas. Nos vemos en Mestalla. **Bona nit!** AMUNT!!!

75. Pues nada, voy a ver Velvet que al menos Miguel Ángel Silvestre me alegra la vista y me endulza la mala leche que tengo!! **Bona nit**

76. Buenas noches// #TwitterOff // #AmuntSempre

77. Allá vamos!!! **Amunt!!!**

78. Se nos viene partidazo! Pase lo que pase: **Amunt Valencia Sempre!**

79. Mestalla ahora! **Amunt Valencia!!!**

80. Confío n éste Valencia, por eso sé que aunq nos lo pondrán difícil, la victoria será nuestra! #amuntvalencia #APorEllos

81. **Bona nit! //Me** voy a soñar con la victoria del domingo. **Amunt!**

82. Felicidades! Eres un currante y te lo mereces! #amunt @valenciacf #JaviFuego2017

83. Ilusión y orgullo. Dos cosas que hacía tiempo no teníamos contigo @valenciacf #SempreAmunt

84. Nos da igual por la mañana, tarde o noche. Mestalla nunca falla! No me falles VCF, yo nunca t he fallado! #Amunt

85. Toca ir a dormir, que mañana hay que levantarse prontito gracias a Tebas. Nos vemos en Mestalla. **Bona nit!** AMUNT!!!

86. 1 de Marzo. // Valencianos, ya estamos en el mes de las Fallas!!! **Vixca Valencia !!!**

87. Cómo me gustan los días de partido y esas ganas que siempre hay que llegar la hora! Cdo juega el VCF el día es diferente! #Amunt #cordobaVCF

88. Olé Quique en rueda de prensa, diciendo que es valencianista y que a ver si queda claro ya! **Amunt Quique!! Amunt Valencia!!**

89. Gooooo! Ahora, por Dios, no la cagueis que nos conocemos! Vamos a por el segundo!!! **Amuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu
95. Agradecido de q no se haya acabado el mundo. No podía ser q pasara el año q el VCF va a pelear por la liga... Jos). Bona nit!!
96. Bon día!!! Empiezo a reaccionar tras el palo del martes. //Los Buenos tiempos llegaran! #SempreValencia
97. Esa intensidad era la que se pedía en Copa, hoy sí, buena segunda parte. Amunt Valencia! <3
98. Muy muy duro... pero esto es así, y como dije: cuando caigas, te levanto. Amunt @valenciacf <3
99. Morir por un sentimiento, vivir con una ilusión. Amunt Valencia
100. Noche de fútbol, noche de copa, vamos mi Valencia #AMUNT
101. QUE GRANDES, QUE GRANDES!!! AMUNT AMUNT VALENCIA!!!
   Dormimos a un punto del líder!!!
102. Goooooooooooooooolazo de ANDRÉ GOMES!!!!!! Cómo me gusta este Valencia!!! Si el año pasado estaba enamorada este año ya... #SentimentVCF
103. Que nerviosa estoy!!!!!!! Amunt amunt amunt VALENCIAAAAA!!!!!
104. Este año sí que sí!!!!!!!!! VAMOOOOOOOOOOS VCF!!! AMUNT AMUNT AMUNT
   #sentimentVCF
105. GOOOOOOOOOOOOOLOAZO DE ANDRÉ GOMES!!!!!! Cómo me gusta este Valencia!!! Si el año pasado estaba enamorada este año ya... #SentimentVCF
106. GOOOOOOOOL! Que bueno eres Paquito!!!!!!! GET 0-1 VAL #AMUNT #AMUNT
   #sentimentVCF
107. Ya ha marcado el desnutrido... Es raro que lo diga pero #amuntllevant
108. Que bonito es ser del Valencia. #TORNAREM #AMUNT
109. Impresionante documental de mi @valenciacf en @user. Por fin algo bueno en la tele!!!!! #SENTIMENTVCF
110. Me parece indignante que Nuno no valore la calidad y el trabajo de @user.
   FORÇAJOAO <3
111. Así da gusto empezar el lunes! #TORNEM #SentimentVCF
112. Que tiemble Sevilla que en cuartos nos vamos para allá! AMUNT AFICIÓ
   AMUNT VALENCIA! #TORNEM @user
113. Pero que Nuno no ve que Barragan las pone fuera del campo???? No hay ni un sólo centro que coloquen bien!!!! #forçaJOAO
114. "Y de orgullo llenaste mi tierra, junts tornem por tu escudo i senyera, siempre juntos contigo Valencia y es por eso que muero por tí."
115. Amunt ostia! Estaremos tocados pero jamás hundidos.
116. ¿Resumen del partido de hoy? // INCREÍBLE// AMUNT!
117. Que gran video @user esto es solo el comienzo, VAMOS A DARLO TODO EN CADA PARTIDO, AMUNT! Tequieuuuu
118. Que bonito es empezar la semana con un exámen. Bon día.
119. A tirar de @user para seguir el Córdoba - VCF. ¡AMUUUUNT!
120. Enhorabuena al @user y a @user por el gol. Pasito a pasito. ¡AMUNT!
121. Desearte mucha suerte en tu nueva etapa en el Madrid gracias por darlo todo por nuestro escudo AMUNT @user
122. "@user: @user #FelizSabadoSara vamos vamos #amunt" Guapa!!!
123. Jajajaja! Déjate! Seguiré siendo normal! :)// Bona nit te quiero!!// @user
124. "@user: @user descasa Saritaaaaaa.Bona nit!!"
125. Al final las cosas improvisadas son las que mejor salen! :)// #Bonanit
126. "@user: #FF @user @user //Bon dia locas! "/De vuelta!!! Os quiero!!!
127. "@user: @user Buenos dias!!!!" Bon diaaa
128. A veces hay que plantearse las cosas si hacen más mal que bien. #Bonanit
129. @user: Bona nit a tots. Amunt!" Bonanit te quiero
130. "@user: @user Buenos dias Sara!!!" Bon dia Carlos!!!
131. Sempre patint... Pero 3 puntos importantes y necesarios! Ahora a pensar en el próximo partido! #Amunt
132. Hay días tontos... Y tontos todos los días!! #BonDia #FelizSábado
133. Comença el partit!! #Amunt Me ha fallado Nuno no poniendo a @user
135. Buenas noches y amunt Valencia.
136. Valencia cf es de champions: https://youtu.be/dr1ObADpFFY Gente que ganas de esto. Ilusión desbordada/#Amunt @valenciacf!!
138. Un aplauso a todos los aficionados que fueron ayer a Elche.Muy grandes.Amunt!!
139. Buenos dias hoy los 96 años del club mas grande que puede existir. Amunt Valencia.
140. Bona nit, orgulloso de mi equipo y de tener amigos como los de @pcvcf_oficial
141. Bona nit y que vuelva pronto la liga, aunque los que marea lo seguirán haciendo.
142. Bona nit que después de intentar ver el vídeo de la "alcaldesa" me han entrado todos los males.
143. Querido @lazaroftherock está foto es mel de romero. Puxa, AMUNT
145. Ahora ya más calmada va AMUNT VALENCIA y a dormir todos que en la cama se esta mejor que en el sofá renegando
146. Mal partido, pero con buen resultado... espero que se estuviesen guardando para Cornellá. Bona nit. AMUNT
147. Ver esto de mi entrenadora del gym y dormir con la conciencia más tranquila JAJAJAJAJA #BonaNit #BuenasNoches
148. Buenas noches #Valencia #goodnight
149. Mal partido, pero con buen resultado... espero que se estuviesen guardando para Cornellá. Bona nit. AMUNT
150. ye pues nada AMUNT VALENCIA! enhorabuena a los del español y continuar patint con el Valencia
151. Muchísimas gracias por este espectáculo, lo habéis dado todo y cada día estoy más orgulloso de ti, te quiero @valenciacf AMUNT
152. BUENOS DÍAS!!!! Hoy es el gran día, la gran batalla en el Camp Nou, AMUNT @valenciacf
153. Amadeo DIOS Salvo cerrando muchas bocas, a ver con quien la pagan ahora los madridistas. A fer la mà!
154. Viendo como está el Barça, nos va a costar mucho sacar algo bueno del Camp Nou. Pero bueno, no hay que rendirse nunca, a por todas, AMUNT
**APPENDIX B: TWITTER INFORMANTS AND NUMBER OF TWEETS PER INFORMANT**

<table>
<thead>
<tr>
<th>Twitter Informant</th>
<th>Number of Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1A3</td>
<td>481</td>
</tr>
<tr>
<td>F2A4</td>
<td>566</td>
</tr>
<tr>
<td>M1G1</td>
<td>116</td>
</tr>
<tr>
<td>M2G2</td>
<td>120</td>
</tr>
<tr>
<td>M3G3</td>
<td>66</td>
</tr>
<tr>
<td>F3A1</td>
<td>29</td>
</tr>
<tr>
<td>F4A2</td>
<td>68</td>
</tr>
<tr>
<td>F5V1</td>
<td>95</td>
</tr>
<tr>
<td>M4A5</td>
<td>79</td>
</tr>
<tr>
<td>M5V2</td>
<td>53</td>
</tr>
<tr>
<td>F6V3</td>
<td>69</td>
</tr>
<tr>
<td>M6C1</td>
<td>30</td>
</tr>
<tr>
<td>M7V4</td>
<td>151</td>
</tr>
<tr>
<td>M8V5</td>
<td>99</td>
</tr>
<tr>
<td>F7V7</td>
<td>116</td>
</tr>
<tr>
<td>M9V9</td>
<td>20</td>
</tr>
<tr>
<td>M10V10</td>
<td>39</td>
</tr>
<tr>
<td>F8G6</td>
<td>67</td>
</tr>
<tr>
<td>M11A6</td>
<td>29</td>
</tr>
<tr>
<td>F9V11</td>
<td>120</td>
</tr>
<tr>
<td>F10V12</td>
<td>62</td>
</tr>
<tr>
<td>M12A7</td>
<td>141</td>
</tr>
<tr>
<td>M13V13</td>
<td>20</td>
</tr>
<tr>
<td>M14V14</td>
<td>133</td>
</tr>
<tr>
<td>F11V15</td>
<td>37</td>
</tr>
<tr>
<td>F12V16</td>
<td>47</td>
</tr>
<tr>
<td>M15V17</td>
<td>47</td>
</tr>
<tr>
<td>F13V18</td>
<td>188</td>
</tr>
<tr>
<td>F14V19</td>
<td>31</td>
</tr>
<tr>
<td>M16V20</td>
<td>194</td>
</tr>
<tr>
<td>M17P1</td>
<td>487</td>
</tr>
<tr>
<td>M18V21</td>
<td>68</td>
</tr>
<tr>
<td>F15V22</td>
<td>193</td>
</tr>
<tr>
<td>F16V23</td>
<td>286</td>
</tr>
<tr>
<td>M19V24</td>
<td>469</td>
</tr>
<tr>
<td>F17Ch1</td>
<td>121</td>
</tr>
<tr>
<td>M20V25</td>
<td>117</td>
</tr>
<tr>
<td>M21V26</td>
<td>18</td>
</tr>
<tr>
<td>M22V27</td>
<td>8</td>
</tr>
<tr>
<td>F18V28</td>
<td>141</td>
</tr>
<tr>
<td>Code</td>
<td>Value</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>F19V29</td>
<td>184</td>
</tr>
<tr>
<td>F20V30</td>
<td>138</td>
</tr>
<tr>
<td>M23G7</td>
<td>425</td>
</tr>
<tr>
<td>F21V31</td>
<td>232</td>
</tr>
<tr>
<td>M24V32</td>
<td>123</td>
</tr>
<tr>
<td>M25V33</td>
<td>123</td>
</tr>
<tr>
<td>M26V34</td>
<td>149</td>
</tr>
<tr>
<td>M27V35</td>
<td>50</td>
</tr>
<tr>
<td>M28O1</td>
<td>40</td>
</tr>
<tr>
<td>M29A8</td>
<td>54</td>
</tr>
<tr>
<td>M30V36</td>
<td>3</td>
</tr>
<tr>
<td>M31V37</td>
<td>86</td>
</tr>
<tr>
<td>M32V38</td>
<td>14</td>
</tr>
<tr>
<td>M33V39</td>
<td>55</td>
</tr>
<tr>
<td>M34V40</td>
<td>174</td>
</tr>
<tr>
<td>M35V41</td>
<td>453</td>
</tr>
<tr>
<td>F22V42</td>
<td>24</td>
</tr>
<tr>
<td>M36V42</td>
<td>290</td>
</tr>
<tr>
<td>M37V43</td>
<td>10</td>
</tr>
<tr>
<td>F23V44</td>
<td>300</td>
</tr>
</tbody>
</table>