

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

### **Flagging in English-Italian code-switching**

Rosignoli, Alberto

*Award date:*  
2011

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **FLAGGING IN ENGLISH- ITALIAN CODE-SWITCHING**

Alberto Rosignoli

Ysgol Ieithyddiaeth ac Iaith Saesneg - School of Linguistics & English Language

Prifysgol Bangor University

PhD 2011



## Summary

This thesis investigates the phenomenon of flagging in code-switching. The term 'flagging' is normally used to describe a series of discourse phenomena occurring in the environment of a switch. In spite of a large literature on code-switching, not much is known about flagging, aside from the general assumption that it has signalling value and may draw attention to the switch (Poplack, 1988). The present study aims to offer a more eclectic understanding of flagging, by looking at the phenomenon from both a structural and an interpretive perspective. The analyses are based on two small corpora of naturalistic conversations collected amongst pairs of bilingual English-Italian speakers in the UK and Italy.

The structural analysis looks at quantitative patterns of flagging. A relationship is observed between the frequency of an item in the data and its production with or without flagging. Higher frequency generally is related to less flagging. A similar relation holds between flagging and different grammatical categories, with nouns being less flagged than adjectives or verbs.

The interpretive approach adopts the methods of Conversation Analysis (Auer, 1998) and investigates how the presence of flagging is instrumental in reconstructing participants' own understanding of the interaction. Through flagging, participants reveal to one another their orientation to single instances of language alternation as belonging or not to the medium (Gafaranga, 2000) of the conversation.

While it may be seen as a peripheral occurrence, flagging can reveal the degree and ease of integration of switches in speech; an appreciation of its role can further the understanding of the dynamics of language contact in naturalistic settings.

By looking at the results from the two analyses, this study shows how flagging is a patterned phenomenon that speakers interpret as having communicative value, rather than a simple disfluency typical of spontaneous speech. ¶



*Does any way less radical exist  
To keep ideals from being trivialised?  
The only way I know is to resist:  
Autonomy cannot be compromised!*

David Hirson, *La Bête*



# CONTENTS

Summary	III
Contents	VI
List of figures and tables	XI
Acknowledgements	XII
Author's declaration	XIII
Abbreviations and symbols	XV
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 <b>Thesis structure</b>	<b>3</b>
<b>2 CODE-SWITCHING: GRAMMATICAL AND INTERPRETIVE PERSPECTIVES</b>	<b>6</b>
2.1 <b>The rise of research on code-switching</b>	<b>7</b>
2.2 <b>Definitions of code-switching</b>	<b>8</b>
2.2.1 Definitions in grammatical studies	10
2.2.1.1 Intrasentential and intersentential code-switching	12
2.2.2 Definitions in interpretive studies	13
2.3 <b>Theoretical approaches to code-switching</b>	<b>16</b>
2.3.1 Structural/grammatical approaches	17
2.3.1.1 The constraints tradition	17
2.3.1.2 The Null Theory of code-switching	18
2.3.1.3 A typological approach to code-switching	19
2.3.2 The Conversation Analysis approach	21
2.3.2.1 What is Conversation Analysis?	22
2.3.2.2 Conversation Analysis and code-switching	23
2.3.2.3 Some principles of a Conversation Analysis-type approach to code-switching	24
2.3.2.4 A turn-by-turn approach to language alternation	25
2.3.2.5 Code-switching as interactional otherness	28
2.4 <b>What are the codes in code-switching?</b>	<b>29</b>
2.4.1 Languages, codes and speaker categories	30
2.4.2 The issues of order and the base language of the conversation	32
2.4.3 The medium of the conversation	33

<b>2.5 Code-switching involving English and Italian</b>	<b>35</b>
2.5.1 English in contact situations	35
2.5.2 Italian in contact situations	37
2.5.2.1 A typology of Italian in contact situations	39
2.5.2.2 Italian varieties and Italo-Romance dialects	40
2.5.3 Italians in the United Kingdom	43
2.5.4 Previous studies on English-Italian code-switching	45
2.5.4.1 Structural/grammatical studies	46
2.5.4.2 Interpretive studies	46
<b>2.6 Conclusions</b>	<b>49</b>
<b>3 THE STUDY OF FLAGGING</b>	<b>51</b>
<b>3.1 Towards a definition of flagging</b>	<b>52</b>
3.1.1 A typology of flagging	56
<b>3.2 Previous work on flagging in code-switching</b>	<b>59</b>
3.2.1 Flagging as indicative of different code-switching styles	60
3.2.2 Flagging as an indicator of discourse-level integration	61
3.2.3 Flagging as indicative of different code-switching strategies	65
3.2.4 Flagging as indicator of a borrowing/code-switching distinction?	66
3.2.5 Flagging from a Conversation Analysis perspective	68
<b>3.3 Does flagging exist in monolingual speech outside code-switching?</b>	<b>70</b>
3.3.1 Hesitation and filled pauses in monolingual speech	73
3.3.2 The notion of repair in monolingual conversation	77
<b>3.4 Concluding remarks</b>	<b>81</b>
<b>4 THE DATA</b>	<b>82</b>
<b>4.1 The two data sets</b>	<b>83</b>
4.1.1 Profile of speakers	83
<b>4.2 Selection of groups</b>	<b>85</b>
4.2.1 Definition of bilinguals	86
4.2.2 Types of groups in code-switching research	87
4.2.3 Groups or communities?	89
4.2.4 Recruitment of speakers	91
4.2.5 Briefing of participants	92

<b>4.3 Ethical issues and speaker consent</b>	<b>94</b>
<b>4.4 Recording procedure</b>	<b>97</b>
4.4.1 Description of recordings	99
<b>4.5 Concluding remarks</b>	<b>107</b>
<b>5 METHODOLOGY</b>	<b>108</b>
<b>5.1 Definition of flagging in the present work</b>	<b>108</b>
5.1.1 Paralinguistic markers	109
5.1.2 Unfilled pauses	110
5.1.3 Lexicalised filled pauses	111
5.1.4 Non-lexicalised filled pauses	111
5.1.5 Other-language equivalents	112
5.1.6 Explicit pre-empting (pre switch only)	113
5.1.7 Explicit justification (post switch only)	113
5.1.8 Disfluencies, false starts, retracing	114
5.1.9 Vowel/syllable lengthening	114
<b>5.2 Data transcription</b>	<b>115</b>
5.2.1 General principles for transcription of the data	116
5.2.2 Transcription system	117
<b>5.3 Features of the CHAT system and transcribing conventions</b>	<b>119</b>
5.3.1 The body of the transcript	120
5.3.2 Language tagging and ambiguous items	121
5.3.2.1 The use of phonological criteria in ambiguous cases	125
5.3.3 Interactional markers	126
5.3.4 Pseudonyms	126
5.3.5 Other transcription conventions	127
5.3.5.1 Pauses	127
5.3.5.2 Spoken and colloquial forms	128
5.3.5.3 Non-standard and regional words	129
5.3.5.4 Overlapping speech	130
5.3.5.5 Unintelligible words/sections	131
5.3.5.6 Vowel and syllable lengthening	131
<b>5.4 Concluding remarks</b>	<b>132</b>

<b>6 DESCRIPTION OF SWITCHES AND QUANTITATIVE ANALYSIS OF FLAGGING</b>	<b>133</b>
6.1 <b>Method of analysis</b>	<b>133</b>
6.1.1 Types of switches in the data	134
6.2 <b>Types of insertions</b>	<b>136</b>
6.2.1 Noun phrases and lone nouns	139
6.2.1.1 Complex noun phrase insertion	140
6.2.2 Adjective insertions	142
6.2.3 Verb insertions	145
6.3 <b>Analysis of flagging</b>	<b>149</b>
6.3.1 Flagging of single nouns	152
6.3.1.1 Establishing a baseline for flagging of nouns	153
6.3.1.2 Quantitative differences	154
6.3.1.3 Differences in types of flagging	156
6.3.1.4 Flagging and frequency: first analysis	159
6.3.1.5 Flagging and frequency: reclassifying the data	161
6.3.2 Flagging of multiword noun phrases	162
6.3.2.1 Flagging and frequency	164
6.3.2.2 Types of flagging	166
6.3.3 Flagging of adjectives	167
6.3.3.1 Flagging and frequency	169
6.3.4 Flagging of verbs	170
6.4 <b>Concluding remarks</b>	<b>171</b>
<b>7 QUALITATIVE ANALYSIS: A CONVERSATION ANALYTIC PERSPECTIVE</b>	<b>173</b>
7.1 <b>Some limits of clause-by-clause analysis</b>	<b>174</b>
7.2 <b>Applying Conversation Analysis to the data: looking at repair</b>	<b>177</b>
7.2.1 Repair in monolingual contexts	179
7.2.2 Repair in bilingual contexts	183
7.2.3 Deviating from the medium	188
7.3 <b>Shifting the boundaries of the medium</b>	<b>193</b>
7.4 <b>Concluding remarks</b>	<b>198</b>

<b>8 DISCUSSION AND CONCLUSIONS</b>	<b>203</b>
<b>8.1 Code-switching in the data: a summary</b>	<b>204</b>
<b>8.2 Results of the analysis of flagging</b>	<b>205</b>
8.2.1 Similarities between groups	207
8.2.2 Flagging and grammatical category	208
8.2.3 Relation between flagging and frequency	209
<b>8.3 Discussion</b>	<b>210</b>
8.3.1 Flagging in code-switching and related discourse phenomena in monolingual speech	211
8.3.2 Flagging and the code-switching borrowing distinction	213
8.3.3 Differences between grammatical categories	217
8.3.4 A note about qualitative analysis	220
<b>8.4 Considerations on the methodology</b>	<b>221</b>
<b>8.5 Suggestions for further research</b>	<b>226</b>
8.5.1 A larger data set	226
8.5.2 Additional types of flagging in naturalistic data	227
8.5.3 The role of frequency vs. categorical differences	228
8.5.4 Flagging as a psycholinguistic phenomenon	229
<b>8.6 Conclusions</b>	<b>230</b>
<b>Appendix 1</b> – Participant contact letter (bilingual version)	<b>233</b>
<b>Appendix 2</b> – Speakers’ social networks	<b>236</b>
<b>Appendix 3</b> – Summary of recordings	<b>237</b>
<b>Appendix 4</b> – Participant questionnaire (bilingual version)	<b>238</b>
<b>Appendix 5</b> – Individual speakers’ questionnaire data	<b>250</b>
<b>Appendix 6</b> – Manchester speakers’ questionnaire data	<b>251</b>
<b>Appendix 7</b> – Milan speakers’ questionnaire data	<b>257</b>
<b>Appendix 8</b> – Sample of a transcript from the Manchester data	<b>264</b>
<b>Appendix 9</b> – Sample of a transcript from the Milan data	<b>275</b>
<b>References</b>	<b>288</b>

## List of figures and tables

<b>Figure 1.</b> Screenshot of a chat transcript with file header and soundwave.	121
<b>Figure 2.</b> Insertions in the two data sets classified by grammatical category.	138
<b>Figure 3.</b> Count of flagged vs. Non-flagged switches in the two data sets.	149
<b>Figure 4.</b> Position and number of flags in Manchester data.	150
<b>Figure 5.</b> Position and number of flags in Milan data.	150
<b>Figure 6.</b> Flagged insertions by position of flags.	152
<b>Figure 7.</b> Rate of flagging for nouns in the two data sets.	153
<b>Figure 8.</b> Flagging of monolingual single nouns.	155
<b>Figure 9.</b> Noun flagging and frequency in Manchester.	159
<b>Figure 10.</b> Noun flagging and frequency in Milan.	159
<b>Figure 11.</b> Noun flagging and frequency in Manchester (recount).	161
<b>Figure 12.</b> Noun flagging and frequency in Milan (recount).	162
<b>Figure 13.</b> Rate of flagging for multiword NPS in the two data sets.	163
<b>Figure 14.</b> Np flagging and frequency in Manchester.	165
<b>Figure 15.</b> Np flagging and frequency in Milan.	165
<b>Figure 16.</b> Flagging rates for adjective in Manchester and Milan.	167
<b>Figure 17.</b> Adjective flagging and frequency in Manchester.	170
<b>Table 1</b> Flagging of switches and borrowings in Jersey French (Jones, 2005)	67
<b>Table 2.</b> Noun tokens and types.	152
<b>Table 3.</b> Types of flagging in non-switched and switched nouns.	157
<b>Table 4.</b> Flagging strategies used for inserted multiword NPS.	166
<b>Table 5.</b> Flagging strategies used for inserted adjectives.	169



## Acknowledgements

This research was partly supported by a Bangor University's College of Arts and Humanities scholarship.

I would like to express my gratitude to my supervisor, Prof Margaret Deuchar, for her constant guidance, support and expertise through all the phases of my PhD.

Special thanks also go to my second supervisor, Dr Maria Carmen Parafita Couto, for always offering a different perspective on my work.

I would like to thank all members of the ESRC Centre for Research on Bilingualism at Bangor University. Throughout my course, their enthusiasm has been a precious source of inspiration. Thanks also go to all staff, past and present, in the School of Linguistics and English Language for their continuous help.

I would like to thank Dr Diana Carter for her help with statistics and Dr Eva Eppler for her comments on parts of the analysis.

A special mention goes to my colleagues in the PhD room, with whom I have had the chance to discuss matters both scholarly and light-hearted, exchange ideas and preserve sanity at times of high stress.

I would also like to thank Veronica Fernandes and Dr Ludovica Serratrice for their help with recruitment, and Sveva Ciaravolo for “having standards”.

My heartfelt thanks go to my family and friends for their continuous encouragement. They may not think much of flagging or code-switching, but they have always remained by my side.

Finally, I would like to say thank you to David Kelly for his virtually inexhaustible patience, constant support and love.

## Abbreviations and symbols

### GENERAL

CA	Conversation Analysis
CS	code-switching
L1	first language
L2	second language
LFP	lexicalised filled pause
NLFP	non-lexicalised filled pause
NP	noun phrase
UP	unfilled pause

### SYMBOLS USED IN TRANSCRIBED EXAMPLES

@it	Italian word/morpheme (in English context)
@en	English word/morpheme (in Italian context)
@0	ambiguous word
@s:en+it	composite word: first morpheme English, second Italian
&	broad phonetic transcription follows
&=	paralinguistic information follows
[>]	overlap begins
[<]	overlap ends
#	pause
[/]	retracing with no change
[//]	retracing with minor change
[///]	retracing with major change
+/"	quotation follows
+"	reported speech
+/.	interruption
+//.	self-interruption
+...	trailing off
<>	scope marking for retracing or overlaps
[?]	uncertain transcription
( )	used to indicate unpronounced part of word

:	vowel or syllable lengthening
. - ? - !	utterance terminators
*ABC	each speaker line is identified by a three letter code preceded by an asterisk

### **SYMBOLS USED IN GLOSSES**

1	first person
2	second person
3	third person
CONDIT	conditional
DET	determiner (article)
FUT	future tense
GER	gerund
IM	interactional marker
INF	infinitive
IMF	imperfect tense
IMP	impersonal particle
OBJ	object
PL	plural
PRN	pronoun
PRS	present tense
PRT	particle
PST	past tense
PTCP	participle
REFL	reflexive / reciprocal particle
REL	relative
SG	singular



# INTRODUCTION

The aim of this thesis is to offer a critical appraisal of flagging in code switching.

The terms ‘flagging’ and ‘flagged code-switching’ were first used in Poplack (1985) when contrasting code-switching patterns in two communities. In one group switches are smooth or ‘skilled’, and speakers seem generally unaware of the occurrence of language alternation, while in the other switches are accompanied by discourse phenomena ‘flagging’ them or showing speakers’ awareness that a switch is taking place.

As its name implies, flagging is commonly understood in the literature as a phenomenon that by its occurrence draws attention to the switch in the proximity of which it occurs. Flagging seems to be quite frequent in some bilingual groups of speakers, but appears to be absent in others (see Poplack, 1988; Eppler, 1994). A variety of discourse phenomena are included within the scope of the term “flagging”. Pauses, hesitation markers and false starts are perhaps the better-known instances, but there are also other metalinguistic and paralinguistic phenomena produced by speakers, such as comments, justifications or laughter. In spite of a vast literature existing now on many aspects of code-switching – defined broadly as the use of two languages in the same speech event – flagging is a somewhat understudied phenomenon, and not a great deal is known about it. Attention may be drawn to a switch in a variety of ways, some more explicit than others: it may therefore be difficult to consider all of them as manifestations

of the same phenomenon. On the other hand, most of the phenomena that are classified as flagging are not exclusive to bilingual speech, and are routinely found in the speech of monolinguals, bilinguals and language learners.

The fact that the label “flagging” encompasses a variety of very heterogeneous discourse phenomena and that the same phenomena are not exclusive to code-switching may be just some of the reasons that have so far prevented flagging from receiving adequate attention in studies of bilingualism. More crucially, however, flagging may have been disregarded in that it represents an eminently performance phenomenon, therefore falling outside the scope of more formal accounts of code-switching.

Our aim is to redress this imbalance and show how, while it may appear to be a peripheral phenomenon, flagging can reveal a wealth of information about what speakers accomplish in code-switching. By investigating original naturalistic data, we will seek to answer a very general question: what does flagging do? More precisely, in our analysis we will address some specific issues: whether flagging is a patterned phenomenon; whether it is more likely to occur with certain types of switches; whether it differs at all from discourse phenomena typical of spoken language; and what it can reveal about how participants themselves organise and understand their interaction, and the fact of using two languages. The answers to these questions will, in turn, inform our understanding of existing issues in the field of language contact, such as the possibility of distinguishing between different outcomes of contact and the separateness of languages in speakers’ repertoires.

In order to study flagging, we will focus on two original corpora of spontaneous English-Italian conversations collected in Great Britain (Manchester) and Italy (Milan) amongst bilingual speakers of the two languages. These will be analysed using both quantitative methods (based on the work of Poplack and associates) and qualitative methods following a conversation-analytic methodology. Using two radically different theoretical orientations and carrying out two separate analyses will result in a better understanding of flagging.

Our primary interest in this study is on functions and features of flagging – and more indirectly code-switching – in two groups which speak the same two

languages but differ with respect to their dominant language. We investigate whether the groups differ substantially in their switching patterns and – closely related to this, the flagging strategies they adopt and the extent to which they use them.

Speakers in both groups are L1 speakers of a minority language in the context of the host country in which they live. Italian as a minority language in English-speaking areas has been widely documented. English as the language of a minority group has not been widely examined outside colonial contexts in which it represented the “high” or prestigious variety, in situations in which language contact was relatively limited.

As it is closely related to the mechanisms of speech production and to disruptions of the speech flow, flagging could be fruitfully studied as a psycholinguistic phenomenon (see Hlavac’s (2011) study of flagging as “hesitation and monitoring phenomena”). This dimension of flagging, however, will not be analysed in detail in the present work, our focus being on the more structural and organisational aspects of the phenomenon.

## **1.1 Thesis structure**

After this introduction, chapter two presents the phenomenon of code-switching as an area of research in linguistics and related fields. The rise of code-switching as a central focus of interest in bilingualism studies is presented, as are some of the main theoretical problems that have dominated the field. The chapter also introduces the distinction between perspectives of code-switching that we are labelling ‘grammatical’ and ‘interpretive’ respectively. This distinction is important in order to frame the two analyses of flagging that will be carried out in this study. While the quantitative analysis exemplifies grammatical approaches, the qualitative analysis represents a more interpretive perspective. The final part of the chapter looks at both English and Italian in contact situations, with a special focus on code-switching. Chapter three focuses on the phenomenon of flagging. The definition of flagging is discussed. Existing work on flagging is also presented, showing how the phenomenon has only occasionally been studied in work on code-switching, and has been used mostly as a diagnostic tool to

characterise different language mixing patterns or distinguish between language contact phenomena. The relationship existing between flagging, hesitation and repair phenomena in monolingual speech – which we indicate as one of the reasons behind the lack of adequate interest for flagging – is also explored. Chapter four introduces the Manchester and Milan data. The features of the groups of speakers are then presented based on questionnaire results. Participants are briefly compared with other types of groups investigated in CS research. The procedures for participants' recruitment and data collection are introduced next, and are followed by a brief description of each of the recordings (participants, setting, and other extra-linguistic information). The methodology for data coding and transcription is the subject of chapter five. The chapter presents a typology of flagging phenomena encountered in our English-Italian data and introduces the working definition of flagging employed in the analyses. The second part of the chapter introduces the issue of transcription and details the methods and conventions adopted for transcription of the recordings. Chapters six and seven constitute the analysis of flagging proper. Chapter six presents the results from the quantitative analysis. The patterns of code-switching are presented before patterns of flagging, since the latter are necessarily related to the former. Insertional code-switching is the prevalent pattern in the data, and the one that the analysis of flagging focuses on. Flagging is analysed according to the grammatical category of the switches and their frequency of insertions in the data. The Manchester and Milan data are presented side-by-side, and differences and similarities between the groups are highlighted throughout. The picture that emerges is one of overall similarities between the groups, but there are some important differences which are highlighted. A more qualitative analysis of the data is the object of chapter seven. The analysis is eminently interpretive, and adopts a conversation-analytic methodology. Given the overall similarities between the groups in terms of flagging strategies, the analysis is conducted on the two corpora treated together. The analysis focuses on the occurrences of flagging in conjunction with instances of conversational repair. This context is particularly revealing of the procedures speakers use to conduct their conversation in an orderly manner.



The occurrence of flagging and repair is shown to be of crucial importance in order to reconstruct participants' understanding of the interaction and categorisations with reference to the languages used in an episode.

Finally, chapter eight presents a summary of the results from the two previous chapters and discusses their implications for existing areas of discussions within the field of code-switching. Suggestions for further research are also presented, both as an extension of the work carried out and as an expansion into related disciplines. ¶

## **CODE-SWITCHING: GRAMMATICAL AND INTERPRETIVE PERSPECTIVES**

**T**his chapter presents a selective overview of research on code-switching (hereafter CS). After a short introduction dealing with the rise of CS as a central subject of investigation in studies of bilingualism, the definition of CS is discussed, with reference to both quantitative and qualitative perspectives. Following this, previous work on CS is reviewed. Particular attention in the review is given to grammatical approaches and to more interpretive approaches – namely work conducted adopting a Conversation-analytic perspective. As we shall see, while grammatical approaches focus on the emergence of regularities in patterns of CS, interpretive approaches tend to concentrate on the discourse uses of CS, participants' own understanding of the interaction and the functions fulfilled by CS. The dichotomy between the two perspectives, however, is not absolute. Related to both approaches is the more general theoretical issue of what the 'codes' in CS are and how and whether they can be distinguished. The question is of particular importance for interpretive approaches, but is also relevant to grammatical studies – especially when CS is considered together with other manifestations of language contact. Since this thesis deals with English-Italian data, a review is given of Italian and English in CS contexts in the final part of the chapter. Particular attention is given to the presence of Italians in Anglophone countries and previous studies on English-Italian CS.

## 2.1 The rise of research on code-switching

Within the area of bilingualism, few topics have been the object of a surge of interest as dramatic as that observed for CS in the last quarter of the twentieth century. In the not too distant past, CS was still considered to be an aberration, a quirk or, at best, the product of an imperfect acquisition of one or more languages in the speakers' repertoire. The very existence of the phenomenon was all but denied. One early study on bilingualism observes that

**except in abnormal cases speakers have not been observed to draw freely from two languages at once. They may switch rapidly from one to another, but at any given moment they are speaking only one, even when they resort to the other for assistance. The introduction of elements from one language into the other means merely an alternation of the second language, not a mixture of the two.** (Haugen, 1950: 211)

In a similar spirit, Weinreich (1953: 73) speaks of the existence of the "ideal bilingual speaker" as someone who "switches from one language to the other according to appropriate changes in the speech situation (interlocutor, topics, etc.), but not in an unchanged speech situation, and certainly not within a single sentence", therefore assuming the existence of normative and degenerate manifestations of bilingualism, and unambiguously relegating CS to the latter category.

A negative or at least ambiguous characterisation of CS remained dominant in the field for several decades; twenty years after the words of Weinreich Labov (1971: 457), while not regarding CS as an aberration, still considered it to be an essentially haphazard and unconstrained phenomenon. In his own words "no one has been able to show that such rapid alternation is governed by any systematic rules or constraints". Patterns of use of the two languages in the same speech event were regarded as simply unpredictable.

A shift in attitudes was evident in the same decade in the work of Blom and Gumperz (1972) and later Gumperz (1982), in which CS was shown not to be a random incident, but to fulfil a number of discourse strategies in conversation. Ever since, interest in the phenomenon has exploded, and CS has gone from

being a peripheral, almost accidental occurrence to one of the central areas of interest within the field of bilingualism and language contact. Today, CS is one of the dominant topics in the field of language contact, and has been subject to analysis from virtually all sub-disciplines within linguistics. Many scholars agree in regarding CS as a 'window' into the nature of language and language contact, as well as the social and cognitive mechanisms underlying speech production and use (Bullock and Toribio, 2009). As a type of language behaviour, CS in some form characterises the majority of speakers around the world, rather than some isolated groups. The rise of interest of CS represents the acknowledgement of its central role in our investigation and understanding of language.

## 2.2 Definitions of code-switching

Considering the prominent role that CS has assumed in recent decades, both in linguistics and neighbouring disciplines, and the proliferation of approaches for its study, it is unsurprising that the ways in which the term 'code-switching' is employed can vary between works, sometimes considerably. Differences amongst scholars include both the issue of what CS actually implies at the level of interaction between languages and the range of phenomena that should be counted as part of the definition. Different uses and scopes depend on authors' theoretical orientation, the particular perspective or discipline from which they are studying the phenomenon and their ultimate research objectives, not to mention the features of the data they work on. Crucially, a significant difference exists between approaches that have a more structural or grammatical focus, and those which attempt to reconstruct participants' perspective and (implicit) understanding of verbal interaction by focusing on the functions of CS in discourse. Broadly speaking, grammatical studies focus on the syntax of CS, while interpretive studies are more interested in its pragmatics. For the latter, the very existence of two discrete languages in CS is a matter of discussion.

The label used to identify the phenomenon may have played a part in this proliferation of definitions. Gardner-Chloros (2009: 11) observes how "both halves of the term CS *are* misleading." She notes how the term originated in the field of

communication technology in the fifties and was only later applied to language interaction (see also Álvarez-Cáccamo, 1998). Both argue that the two concepts of ‘code’ (which has been read as shorthand for “language variety”) and ‘switching’ (which underlines the instantaneous and complete activation/deactivation of either code) have perhaps contributed to generate some confusion.

With reference to the scope of the definition, Poplack (2004: 28) remarks how “[t]here is little consensus in the literature over which aspects should be subsumed under the label code-switching.” Clyne (2003: 70) notes, for instance, how it is possible to distinguish in the literature between those that use the term CS in opposition to borrowing and those that use it as subsuming borrowing. Some scholars prefer to use other labels – such as “code-mixing” (Muysken, 2000) or “language alternation” (Auer, 1998) instead of “code-switching” to describe the use of two languages in the same speech event. When alternative terminologies are proposed, they do not generally displace the label CS altogether. Typically the term “code-switching” is still used by proponents of alternative terminologies, but is reserved for some more specific manifestations of language contact phenomena. This is particularly, but not exclusively, the case for work conducted from a Conversation Analysis perspective (see sections **2.2.2**; **2.3.2**).

In spite of these varying preferences, CS still appears to be the more commonly used term by students of bilingualism to describe the use of more than one language in a given speech event (Clyne, 2003: 71). The same practice will be adopted here. In what follows, several definitions of CS are presented and discussed, with particular attention to the similarities and differences between them and their varying scope. Both definitions used in grammatical studies and definitions typical of interpretive work (particularly conversation analytic work) will be outlined.

One feature that will be highlighted is that while in most studies on the grammar and structural aspects of CS a substantial equivalence holds between the notions of ‘code’ and ‘language’ to indicate the languages or language varieties involved in CS, this is not necessarily the case in studies adopting a conversation analytic approach. The non-equivalence of the two notions brings about a re-specification of the ‘codes’ alternating in conversation, and a redefinition of the very idea of ‘code-switching’. This in turn may involve the use of alternative terminology.

### 2.2.1 Definitions in grammatical studies

Poplack (1980 [2000]: 224) defines CS as “the alternation of two languages within a single discourse, sentence or constituent” and as “the juxtaposition of sentences or sentence fragments, each of which is internally consistent with the morphological and syntactic (and optionally, phonological) rules of the language of its provenance” (Poplack, 1993, in Muysken, 2000: 14). CS is therefore characterised as a rather broad phenomenon; the alternation of the two languages may occur at several levels in a speaker’s production: within a conversation, within a sentence or within a constituent.

A rather different characterisation is offered by Backus, who defines CS as “any kind of discourse in which words originating in two different language systems are used side-by-side” (2005: 307). Backus’ definition encompasses a wide range of possible contact situations, most notably including lexical borrowing. CS and borrowing are treated unitarily as different manifestations of the same phenomenon, within the general perspective of contact-induced language change. While CS is typically looked at from a synchronic perspective, and studies on borrowing normally adopt a diachronic point of view, Backus argues, they are essentially the two sides of the coin. In Backus’ own words “insertional<sup>1</sup> CS is the synchronic reflex of lexical borrowing” (2005: 315; see also Backus and Dorleijn, 2009: 79).

Similarly to Poplack, Clyne defines CS as “the alternative use of two languages either within a sentence or between sentences” (1987: 258) therefore also stressing how CS may occur at different levels in the discourse. The reference to alternation “within” as opposed to “between” sentences highlights a distinction that characterises much research on CS, i.e. the one between intersentential and intrasentential CS (see 2.2.1.1). For studies investigating grammatical constraints on CS, the distinction between the two is a salient one; as Myers-Scotton (2002: 55) argues, it is only below the level of the sentence – at the level of the projection of complementizer – that the two or more languages involved in a multilingual speech event can be said to be really in contact (see also Joshi, 1985: 190). In more recent years, Clyne, while still accepting the generalised use of “code-switching” proposes

---

<sup>1</sup> Backus (2005: 321) identifies two subtypes of CS, ‘insertion’ and ‘alternation’. For an explanation of the terms, see section 2.3.1.3 about Muysken’s (2000) typology of code-mixing.

in his model of contact the umbrella term “transference” (2003: 72), which is said to cover all types of linguistic elements that can be transferred from one language to the other. Clyne suggests a more comprehensive framework, which includes amongst others morphemic, prosodic and pragmatic transference of elements from the languages in contact. The use of the term “transference”, however, does not seem to have been widely adopted in the literature.

MacSwan (2000: 38) characterises the phenomenon as one of juxtaposition: “Code switching is a speech style in which fluent bilinguals move in and out of two (or more) languages.” No mention is made of the level at which the movement from one language to the other may occur. This absence, however, is not the most remarkable feature of the definition; rather, it is the fact that CS is produced by speakers that are characterised as being “fluent bilinguals”, that is to say, having a high degree of proficiency in all of the languages involved in the conversation. This view contrasts with the one proposed by Poplack (1980 [2000]) and Myers-Scotton (2002) amongst others, who claim that a high level of proficiency in both languages is not required in order for CS to take place. Various types of CS have been characterised in terms of speakers’ proficiency in the languages involved, since linguistic ability can to an extent affect the kind of CS patterns that speakers produce (Poplack, 1980 [2000]: 231; Finlayson *et al.*, 1998). MacSwan’s definition restricts instances to be considered as legitimate CS based on a very specific criterion, i.e. the speakers being balanced bilinguals. In more recent work, he seems to qualify such a claim; in MacSwan (2009: 309) the definition “CS involves the mixing of phonologically distinctive elements into a single utterance” is adopted instead. The reference to distinct phonological systems still refers to the fact that “true” CS is only produced by simultaneous or at least balanced bilinguals, since the assumption is that only these will have sufficiently separated phonological systems for each of their languages.





A more general problem is represented by the underlying assumption that spoken data can be easily divided into sentences or clauses, which is not always the case when dealing with naturalistic data.

Besides inter and intrasentential, some researchers also identify a third category of “tag” or “extra-sentential” CS (see Poplack *et al.*, 1989; Romaine, 1995), involving elements with no syntactic relation to the rest of a clause or sentence, such as discourse markers (e.g. English *you know, so, thank God*).

As shown by this review of definitions, the distinction between intra and intersentential CS is of some consequence for grammatical approaches to CS, since the sentence represents the upper limit of grammatical analysis. In more interpretive approaches to CS, on the other hand, the distinction is not necessarily made or treated as significant.

### 2.2.2 Definitions in interpretive studies

While in the definitions of CS in grammatical studies the equivalence between languages (or varieties) and codes is taken for granted, more interpretive studies of CS, such as those adopting a Conversation Analysis (CA) approach, reject the equivalence between language and code in CS and have reclassified the concept of CS and its domain accordingly.

The issue is hinted at by Gumperz (1982: 99), who argues that speakers attend to their own notion of communicative code, not necessarily corresponding to any fully-fledged language. This observation, however, does not result in a rejection of CS as an umbrella term for language contact phenomena.

In more recent years, and particularly in authors working from a conversation-analytic perspective, the lack of coincidence between the codes involved in CS and the participating languages, as defined from an external observer’s point of view, has been at the centre of a terminological and conceptual debate.

Emblematic of this current is the work of Peter Auer: while he appears to accept the definition of CS as “the alternating use of two or more “codes” within one conversational episode” (1998: 1), he then goes on to specify that “the question of what counts as a code is not easily answered, for it must refer to participants’, not to linguists’ notion of ‘code A’ and ‘code B’” (1998: 13). One of the tenets of

Auer's approach is the interpretation of speakers' verbal actions through their orientation (response) in adjacent conversational turns. In bilingual conversation, therefore, the question of what counts as a code, upon which the definition of CS necessarily rests, is crucial to an interpretive approach. The question cannot be answered simply by resorting to the idea of language or language variety; these categories – Auer argues – are brought to bear on the data by the analyst and may therefore not be relevant to speakers. In order for an instance of language alternation to be interpreted as CS, analysts need to ascertain whether the alternation is “perceived by *participants* as involving different ‘codes’” (Auer, 1998: 15, italics in the text). The concept of CS therefore undergoes a rather radical re-interpretation and is used to indicate those instances of language alternation that are treated by speakers as signalling some kind of change from the immediately preceding turn. The interpretation of language alternation must be sequential, i.e. on a turn-by-turn basis, since this reconstructs participants' own methods of dealing with the on-going conversation.

In order to do this it is necessary to distinguish between ‘languages’, which are determined externally and ‘communicative codes’ that are recognised by actors in a conversation. Communicative codes are defined as “mechanisms of transduction between intentions and utterances and then between utterances and interpretations” (Álvarez-Cáccamo 1998: 38). Paraphrasing Álvarez-Cáccamo, Gafaranga and Torras (2002: 10) note that in conversation participants rely on a variety of communicative codes, both verbal and non verbal, and that language choice represents only one among such codes.

If the codes in CS are not equivalent to languages, then not all instances of language alternation can be considered to be CS. Also, CS becomes only one of many language-contact phenomena which from an analyst's perspective (as opposed to a participant's) represent the juxtaposition of two language or varieties. One example other than CS is “code-mixing”<sup>3</sup>, a type of speech showing numerous instances of language alternation which individually do not carry meaning but which constitute a speech style or separate variety for some speakers (Auer, 1999).

---

**3** “Code-mixing” here is not to be confused with the use Muysken (2000) makes of the term as the umbrella term for the use of two languages in the same speech event.

A different approach within the field of CA applied to language alternation is offered in the work of Gafaranga (2000; 2007). While indebted to Auer, most notably regarding the adoption of participants' perspective in the interpretation of bilingual interaction, Gafaranga differs from Auer in that he claims that the latter's notions of 'language' and 'code' are in fact too similar to one another. He also appears to agree with Álvarez-Cáccamo's characterisation of 'communicative code'; according to this view, language choice is only one of the communicative codes relied upon by speakers in a conversation. Gafaranga (2000) proposes to rename the particular communicative code concerned with language and language choice as the "medium of a bilingual conversation" in order to highlight its specificity and distinctness from an externally-defined idea of language. Since a medium does not correspond to a language, it can be bilingual as well as monolingual (Gafaranga and Torras, 2002: 12). Concepts similar to that of a bilingual medium are the 'mixed code' in Auer's (1998) terms or 'code-switching as the unmarked choice' in Myers-Scotton's Markedness Model (1993b). The non-equivalence of language and medium entails that not all instances of language alternation are to be treated as CS. Some of these could be part of a bilingual medium, with no particular meaning for speakers, who orient to it as a unitary variety. If this approach is taken, the definition of CS has to be respecified, since not all instances of bilingual speech will have the same status. The term 'code-switching' is then only used by Gafaranga to indicate "any instance of deviance from the current medium which are not oriented to by participants themselves as requiring any repair" (Gafaranga and Torras, 2002: 18). When language alternation itself is the medium, it is not to be considered as CS; in participants' scheme of interpretation, the communicative code – or medium – remains the same, in spite of the observable material from more than one language.

The conversation analytic approach to CS is discussed more in detail in section **2.3.2**.

## 2.3 Theoretical approaches to code-switching

Within the vast literature now existing on CS, it is possible to identify some main strands of research that have developed in the last few decades. There is some disagreement between scholars in the classification of approaches, and different categorisations are used to capture the variety of approaches to the phenomenon.

Tabouret-Keller (1995) classifies approaches according to the constraints on CS that they investigate (linguistic, social or biological constraints). Other researchers (Gardner-Chloros, 2009; Bullock and Toribio, 2009), have proposed more traditional taxonomies. Based on both their classifications it is possible to identify four main strands of research in the field: the structural/grammatical; the psycholinguistic; the sociolinguistic and the pragmatic/conversational.

While grammatical studies are mostly interested in the principles governing the interaction between the two languages, psycholinguistic studies address issues such as the simultaneous activation of the two languages and the cognitive and processing costs of switching. Research carried out within the sociolinguistic and pragmatic/conversational strands tends to adopt both quantitative and qualitative perspectives, and look at general patterns as well as taking a more interpretive stance in some cases. Some conversational work, particularly in Conversation Analysis, could be classified as a more qualitative and eminently interpretive approach to the study of CS. This is because Conversation Analysis adopts an emic perspective (i.e. not referring to an external interpretive framework but attempting to reconstruct speakers' own interpretation and methods for understanding the interaction).

An exhaustive review of work on CS in all the above-mentioned strands is beyond the scope of the present work. For the purposes of the analysis of flagging, I have elected to focus only on structural/grammatical approaches (2.3.1) and conversation analytic studies (2.3.2). The latter, in particular, are taken here to epitomise the interpretive approach to the study of CS.

### 2.3.1 Structural/grammatical approaches

One of the earliest and most influential works on CS within the grammatical tradition is Poplack's study of Spanish-English CS in New York (1980 [2000]). The study had both a grammatical and a more social focus, but it is with the former that we are concerned here. Poplack identifies two principles (the Free Morpheme Constraint and the Equivalence Constraint) supposed to account for the points at which languages may be switched or not. The Free Morpheme Constraint prevents switching between bound morphemes, e.g. word-internally, while the Equivalence constraint indicates switching sites as points of linear equivalence between the two languages.

The study also constituted one of the first attempts to operationally make a distinction between CS on the one hand and lexical borrowing on the other, based on the variation of patterns of lone items in speech. This method has subsequently been developed by Poplack and associates, and is described in detail in Poplack and Meechan (1998).

#### 2.3.1.1 The constraints tradition

The grammatical approach to CS gained considerable momentum with the so-called constraints tradition and research based on naturalistic corpora after Poplack's seminal work. Starting from the assumption that CS is rule-governed and not a haphazard phenomenon, scholars have investigated CS in its naturalistic setting in a variety of sociolinguistic situations and language combinations and outlined constraints on what combinations of the grammars of the languages involved can and cannot occur. Notable examples include Sankoff and Poplack (1981); Joshi (1985); Di Sciullo *et al.* (1986); Belazi *et al.* (1994); Myers-Scotton (1993a; 2002) just to name a few. Proposals differ quite substantially in the types of constraints that they have placed on CS, but they are similar in claiming the validity of their findings beyond the data they investigate. Periodically, however, problematic data involving a new or an already studied language pair is found, which results in the particular constraint being rejected or

re-defined (see MacSwan, 2009 for a review of problematic data for most of the aforementioned studies).

Of particular influence in the literature has been the Matrix Language Frame Model by Myers-Scotton (1993a; 2002) and associates. The model postulates a basic asymmetry between the participating languages in CS. One of the languages (Matrix Language) provides the morpho-syntactic framework of bilingual clauses and most morphemes, while the other language (Embedded Language) contributes only some morphemes, most typically lexical items.

### 2.3.1.2 **The Null Theory of code-switching**

Besides the constraints tradition, within grammatical approaches, several scholars have studied CS patterns without invoking constraints specific to language alternation but attempting to account for CS mechanisms within the general properties of the two languages involved. Emblematic of this work is what MacSwan (2000: 43) refers to as the agenda of the Null Theory approach to CS: “Nothing constrains code switching apart from the requirements of the mixed grammars.” CS is therefore seen by MacSwan as the union of grammatical requirements of the participating languages, with no CS-specific theoretical apparatus. These approaches are characterised by an explicit rejection of CS-specific mechanisms, which are seen as an additional third grammar only relevant in language mixing contexts – and therefore as stipulative and non-economical constructs. Authors arguing for a Null Theory approach to CS mostly come from a generativist background and include Mahootian (1996), MacSwan (2000, 2009) and Chan (2009).

An example of the application of the Null Theory is the work of MacSwan. Based on an interpretation of Chomsky’s Minimalist Programme, MacSwan proposes a lexicalist model in which the syntactic features of lexical items constrain what combinations of the two grammars may occur.

Rather than a well-defined theoretical framework, the Null Theory is described by its proponents as a research programme, which attempts to account for CS within existing syntactic theories. This can give rise to different accounts within

the programme. As noted in Chan (2009: 187) “a major drawback of the Null Theory is that constraints are necessarily expressed in theory-specific terms, and therefore the empirical predictions may vary when researchers are committed to different linguistic theories or assumptions”.

In addition to using naturalistic CS data, scholars in this area have also relied on grammaticality judgements to construct their accounts of CS, consistently with the methods of investigations of generative syntax and an interest in speakers’ competence, rather than aspects of their performance.

### 2.3.1.3 A typological approach to code-switching

Muysken (2000) takes quite a different stance from both the constraints tradition and the Null Theory, with regard to the breadth of his analysis and the nature of patterns extracted from CS data.

Based on the analysis of a large number of bilingual data sets, Muysken identifies three overarching mechanisms of CS<sup>4</sup>, i.e. different ways in which the grammars of the two languages interact in bilingual speech. These mechanisms are insertion, alternation, and congruent lexicalisation.

Insertion is explicitly characterised in terms of asymmetry between the languages, and it involves the embedding of a language B element into a structure (or frame) provided by language A. Muysken acknowledges this pattern to be akin to the predictions of the MLF, even though he sees the concept of the Matrix Language as an essentially empirical/descriptive phenomenon rather than a theoretical construct. Insertional CS is also closely related to lexical borrowing, from which, according to Muysken, it may not always be distinguished.

The second pattern, alternation, presupposes that code-mixing will mostly occur at points where the two languages involved share the same surface linear order. In this, it is certainly reminiscent of Poplack’s Equivalence Principle. One of the main differences from insertion is that the languages “present in the clause remain relatively separate” (2000: 96). In insertion, on the other hand, elements

---

<sup>4</sup> Muysken uses the term “code-mixing” instead of CS as a general label.

from one language are “nested” into the frame provided by the other language. A further difference concerns the length and complexity of the switched elements. While insertion typically involves lone nouns and other lexical items (hence why insertion is considered as non-substantially different from lexical borrowing), alternation usually affects larger elements, or even full sentences. Other elements that do not present strong syntactic integration in the clause, such as tags or discourse markers, are also considered as instances of alternation (2000: 99). In many corpora, alternation seems to occur predominantly between sentences rather than within sentences, as noted in Deuchar *et al.* (2007: 336). The third type of mixing, congruent lexicalisation, partly differs from the other two, inasmuch as it presupposes a certain degree of similarity between the lexicons and the grammars of the languages involved. When some structures are shared between the two languages, these will be realised with elements drawn from both. Two factors facilitating the occurrence of this pattern are the presence of a large number of homophonous diamorphs<sup>5</sup> acting as triggers and a certain degree of structural equivalence between the languages, both in terms of linearisation and grammatical categories (Muysken, 2000: 123). Congruent lexicalisation is characterised as being somewhat related to style-shifting in monolinguals (Muysken, 2000: 68).

Muysken’s interest is mostly taxonomic and classificatory, rather than strictly predictive of possible patterns in CS. This sets him apart from the grammatical work reviewed so far. Muysken explicitly attempts to relate typological linguistic factors and wider community dynamics as facilitating or inhibiting forces contributing to shape the exact dimension of interaction between two or more languages. Also, and perhaps more crucially, the three processes of CS which he identifies are not seen as absolute constraints, as in most of the works mentioned earlier, but rather as general trends to be found in the data. Aside from taking into account very heterogeneous data from a large number of bilingual groups, the significance and originality of Muysken’s work lie in the reinterpretation of absolute CS constraints typical of grammatical approaches in a more probabilistic

---

<sup>5</sup> By homophonous diamorphs Muysken (2000) means cognate words that have similar or identical phonological realisations in the two languages. This is more likely to occur with languages that are genetically related.



sense. Different CS patterns are seen as the product of a variety of linguistic and extra-linguistic factors, and the resulting constraints need not be absolute.

This represents a more holistic and integrated approach in the search of regularities in CS. Grammatical and structural accounts play a key role in our understanding of CS, yet it should be borne in mind that CS is not produced by speakers in a vacuum, but is situated both interactionally and socially.

While CS constraints per se do not fall within the scope of the present work, a view similar to Muysken's approach will be adopted here for the structural study of flagging. An understanding of the structural dimension – even in a quantitative sense – is essential but not sufficient to achieve a comprehensive appreciation of the phenomenon. A more eclectic approach is advocated which integrates those insights with the findings from a more interpretive framework. As already mentioned, amongst the existing strands of research we shall focus here on the contribution offered by Conversation Analysis, as a representative of interpretive approaches to CS.

### **2.3.2 The Conversation Analysis approach**

Studies falling within the Conversation Analysis (hereafter CA) tradition start from radically different assumptions compared with grammatical studies of CS. As mentioned earlier, one of the main differences is that they do not automatically postulate that bilingual conversation involves alternation between two externally defined languages, but rather attempt to reconstruct participants' categories in interaction and the methods by which they organise the linguistic resources at their disposal, by creating a contrast that is interpreted as interactionally meaningful. Virtually all work carried out within the grammatical research tradition assumes that the 'codes' that speakers alternate between in CS correspond exactly to the two languages in conversation. This view, however, does not take into account of the ways in which speakers organise their repertoires and use them in interaction. The equivalence between the languages (seen by the external observer) and the codes in CS has been problematised and called into question at least since Gumperz's (1982) work on CS as a discourse

strategy, in which it is claimed that participants attend to their own notion of code. The legacy of Gumperz's research has been a body of work focussing on pragmatic/interactional accounts of CS, particularly from a CA perspective.

### 2.3.2.1 What is Conversation Analysis?

Conversation Analysis was established as a research methodology in seminal works by Sacks and associates (Sacks *et al.*, 1974; Schegloff *et al.*, 1977). CA has its roots in the field of sociology, particularly in the tradition of ethnomethodology (Garfinkel, 1967), a method of enquiry attempting to reconstruct the methods participants use to arrive at an understanding of interactions in everyday conversations. CA is not a fully-fledged theory of language or interaction, but rather a “mentality” (Gafaranga, 2009: 115), i.e. a way of approaching the study of conversation. CA “has consistently resisted theorizing” (Gafaranga, 2009: 115), something that could be seen as a possible weakness of this method of investigation. Consistently with the ethnomethodological roots of this approach, researchers adopting a CA perspective do not typically formulate hypotheses which are then tested in the data, but proceed inductively from the unfocused observation of data (Edwards, 2006) to attempt to reveal rules of interaction that are demonstrably relevant to speakers.

Central to the CA approach is the notion of ‘sequentiality’, i.e. the idea that talk is organised in a step-by-step fashion. It is argued that the meaning of any utterance cannot be interpreted in a vacuum, but needs to refer back to the immediately preceding context. This represents the framework within which participants’ verbal acts can be interpreted and which informs the structure and content of the ensuing talk. The notion of interpretation, however, is based on what can be seen as relevant for speakers themselves, rather than for the analyst. In order to reveal speakers’ organisation of the interaction, CA relies typically on qualitative rather than strictly quantitative evidence. The “single case” (Schegloff, 2010) is of primary importance to the analyst adopting a CA perspective, since it represents an instantiation of the methods used by speakers to deal with the general organisation of talk. Schegloff stresses the importance of the single case in

CA work and its key role for the analysis of data. While speakers may be applying general organisational principles of interactions, the only way for analysts to discover them is to look at single instantiations of those principles in actual talk:

*Speakers construct the talk via the generic organizations of practice as embodied in that moment's detail, recipients analyze the talk via the generic organizations of practice as embodied in that moment's detail, speakers orient to the recipients doing so, and recipients orient to the speakers having done so.* (Schegloff, 2010: 134, italics in the text)

The importance accorded to single occurrences of a given phenomenon used as evidence contrasts with much variationist or more grammatical work on language and indeed CS, in which single occurrences, especially if deviant, are dismissed as speech errors or as idiosyncrasies. Deviant cases can be of critical importance in CA; it is by analysing how speakers react to (apparent) breaches of a conversational rule that analysts can show speakers' normative orientation to the rule itself (Edwards, 2006).

### 2.3.2.2 **Conversation Analysis and code-switching**

CA was born as a contribution to the field of sociology and with a focus on monolingual language, namely on English. The monolingual focus is still predominant in much of the work conducted by Schegloff and associates, although other languages have been also analysed in recent decades by other researchers. A conversation-analytic methodology was first applied to CS by Auer (1984, 1998) with reference to Italian-German and Spanish-German data in Switzerland. Subsequent analyses carried out following a CA methodology include work by Li Wei (1994) on Chinese-English in the United Kingdom and Gafaranga (2000) on Kinyarwanda-French data in Belgium. An edited volume by Auer (1998) comprises contributions applying Auer's approach to a variety of language contact situations.

The adoption of CA methodology by Auer started as a move forward from Gumperz's (1982) work on CS as a discourse strategy to account for the use of two

languages in a single conversation. More generally, Auer was dissatisfied with macro-level sociolinguistic accounts of CS, such as Fishman's (1965) notion of language domains, which Auer saw as too mechanistic (see Li Wei, 1994 for a review). Auer (1998: 3) writes that "macro-sociolinguistic investigations of code-switching restrict themselves to analysing the social meaning of the occurrence or non-occurrence of code-switching in an interaction at large, but they fail to account for [...] local processes [...] either because they dismiss the question itself as uninteresting, or because they do not believe that such an account is possible". Conversation analytic work represents an intermediate level of analysis between micro-level grammatical constraints – which are mostly concerned with CS at the intrasentential level – and macro-level sociolinguistic explanations (Alfonzetti, 1998: 180).

The application of CA methods to CS has become established as a partially separate research tradition from the original approach developed by Sacks and associates. If the two approaches are compared, it becomes apparent that only some features of CA have been adopted in the study of bilingual conversation. In particular, from a methodological standpoint, the level of detail at which the data are encoded tends to be somewhat lower in CS studies than in traditional CA. When talking about CA, unless otherwise specified, we shall be referring to the adaptation of the CA method to bilingual speech.

### **2.3.2.3 Some principles of a Conversation Analysis-type approach to code-switching**

According to Alfonzetti (1998), in bilingual conversation speakers can use the languages or varieties at their disposal as communicative resources, without necessarily always referring to indexical values attributed to the languages by the community at large, as investigated in sociolinguistic approaches to CS.

The "we-code vs. they-code" opposition in Gumperz's work, for instance, may not always be referred to by participants when engaging in CS. While these values are sometimes necessary for the understanding of CS, the meanings of the languages and of their alternation are negotiated locally by speakers, i.e. within the conversational episode.

Participants' observable verbal actions and reactions to each other's speech are

the evidence through which the analyst can attempt to reconstruct the categories used by speakers. In some contexts, it is the act of switching itself, at a certain point in the conversation, which is interpreted as significant by participants, irrespective of the particular symbolic values externally associated with one or the other language or the direction of the switch (see example **3** below). Meeuwis and Blommaert (1998) argue that in some contexts, CS may be seen by speakers as a variety in its own right, and other parameters within the repertoire, such as density of mixing or syntactic structures, may be seen as more meaningful than the language of particular utterances. Speakers can thus interpret what the external observer would call language alternation as a unitary lect.

As noted when discussing definitions of CS in section **2.2**, therefore, in this strand of research, not all instances of language alternation should be seen as CS; the term is normally reserved for those cases that can be shown to have been used by speakers meaningfully in the interaction, or seen as instances of “interactional otherness” (Gafaranga and Torras, 2002). Adopting a conversation-analytic mentality therefore entails asking the question “what are the ‘codes’ in ‘code-switching?’” (Auer, 1998: 2).

#### **2.3.2.4 A turn-by-turn approach to language alternation**

Auer draws on Gumperz’s (1982) idea of CS as a ‘contextualisation cue’, defined as “activities by participants which make relevant/maintain/revise/cancel some aspects of context which, in turn, is responsible for the interpretation of an utterance in its particular locus of occurrence” (Auer, 1995: 123). As a contextualisation cue, language alternation may not have a meaning outside of the context in which it occurs; rather, its interpretation depends on its locus of occurrence or, in the case at hand, the sequential environment represented by the immediately preceding turns in conversation. The interpretation of code-alternation as a cue is therefore also strongly related to patterns of language choice (Auer, 1995: 124).

Auer distinguishes between two types of language alternation: depending on the role they are interpreted as having in the interaction, alternations can be either

“discourse-related” or “participant-related”. Discourse-related alternation refers to any instance which contributes to structuring the interaction (e.g. the repetition of an item in another language for emphasis, addressee selection, framing of reported speech). In (3) is an example of discourse-related alternation between Sicilian (in italics) and Italian (in plain script)<sup>6</sup>. The speaker switches from Italian to Sicilian to signal that she is changing the topic of the conversation, after which she returns to Italian.

**3.** *A phone conversation. The speaker is talking about politics, then changes topic, and asks about a friend she has been unable to contact.*

Ma l’hai visto? Io mai l’ho vista una campagna elettorale così. Questo dicevamo oggi con A. Neppure nel quarantotto, che era il dopoguerra, che c’erano... che c’erano proprio umori tremendi. Mai s’era verificato. *N’altra cosa t’ai’a cchièdirti*, G. Cambiamo discorso. Io continuo a telefonare a M. Perché è da Pasqua che le voglio fare gli auguri, le cose. Perché non la trovo?

---

Have you seen it? I’ve never seen an electoral campaign like this one. I was talking about this to A today. Not even in 1948, in the post-war period, when there were... when there were tremendous emotions. It’s never happened. *I’ve got something else to ask you*, G. Let’s change topic. I keep on calling M. I’ve been trying to give her my wishes, and so on, since Easter. Why can’t I find her?

(Alfonzetti, 1998: 198)

Participant-related alternation (also called “preference-related”), on the other hand, is typically motivated by speakers’ or interlocutors’ language competence or language preference. In the example in (4) speaker Ri changes languages and shows her preference for the language in which she is more competent, i.e. English (in italics) over Italian (in plain script).

---

<sup>6</sup> Examples taken from the literature are sometimes transcribed following different conventions to the ones adopted here. Where these differences are not felt to hinder the reader’s understanding of the example itself, they are maintained as closely as possible to the original.

4. *A conversation about school between R (Italian researcher), Ri (daughter) and Vi (father)*<sup>7</sup>.

R: hai fatto le scuole italiane?

Ri: si qui al – al – al consolato come si dice? (0.7) e poi anche al – alla scuola – *can't I speak to you in English?*

R: ehe ehe ehe

Vi: no – ma se hai fatto le scuole italiane parla italiano!

Ri: *I used – I used to go to Italian school = that the council – that the government from Italy organise – and I also did my A level at school – full time school – that's the Italian I have done* ehe ehe ehe

R: ehe ehe ehe okey

Ri: adesso faccio un po' di fatica a parlare.

R: si?

Ri: si, ehe ehe! ecc.

---

R: did you attend Italian schools?

Ri: yes here at the – at the – at the consulate how do you say? (0.7) and then also at the – at the school – *can't I speak to you in English?*

R: [laughs]

Vi: no – but if you've gone to Italian schools speak Italian!

Ri: *I used – I used to go to Italian school = that the council – that the government from Italy organise – and I also did my A level at school – full time school – that's the Italian I have done* [laughs]

R: [laughs] okey

Ri: now I struggle a bit to speak

R: yeah?

Ri: yeah, [laughs]! etc.

(Panese, 1992: 74)

Both discourse and participant-oriented alternation can be instances of either CS or what Auer calls “transfer”. While CS is related to a specific point in the conversation, and entails a changing of the language of interaction (Auer, 2000: 177), transfer represents the simple insertion of an other-language structure in the discourse, without changing the language of interaction (always considered from a turn-by-turn perspectives).

---

<sup>7</sup> The English translation of the example is my own, as it is not provided in the original work.

### 2.3.2.5 Code-switching as interactional otherness

Unlike Auer, who approaches the question of language alternation strictly on a turn-by-turn basis, Gafaranga argues for the idea of an overall base code of the interaction, which he identifies as the medium of the conversation (see section 2.4). He argues that

As an instance of social action, language alternation among bilingual speakers must be assumed to be orderly for, if it were not, it would not occur (Garfinkel, 1967). That is to say, talk among bilingual speakers must be assumed to be “informed” by some scheme of interpretation on the level of language choice, by a “grid” (Heritage, 1984) with respect to which speakers themselves make sense of their language alternation activities. The base code of a bilingual conversation would be that scheme of interpretation. That is, the issue of the base code is not a concern for linguists only. It is also a concern for speakers themselves.

(Gafaranga, 2000: 328)

As already mentioned in section 2.2.2, Gafaranga calls this scheme of interpretation the “medium” of the conversation. Analysts have the task of discovering participants’ scheme of interpretation in order to be able to interpret single instances of bilingual speech. This is done by looking at the activities speakers accomplish in conversation in relation to the language of their utterances. From the participants’ perspective, language alternation itself can function as the medium of the conversation. In that case, speakers do not orient to single instances of alternation as being functional. When language alternation is not the medium, i.e. it is not oriented to as normative by speakers, then an instance of alternation must represent an instance of deviance. In turn, any instance of deviance<sup>8</sup> will be interpreted as either repairable or as functional, i.e. serving a communicative purpose. The presence of repair work by speakers after an instance of language alternation is what warrants its interpretation as a

---

<sup>8</sup> Gafaranga (2007) illustrates the idea of normative/deviance interpretation of social actions with the example of greetings. If a speaker A greets B, B may either 1. return the greeting (observing the norm) or 2. not return the greeting (deviant behaviour). In the absence of a response, speaker A will have to interpret the absence of a reply as either 2a. repairable deviance (“He did not hear me” and repeat the greeting) or as 2b. functional deviance (“He is ignoring me”).



case of repairable deviance. When an instance of deviance is not repaired, it is seen as “interactional otherness” (Gafaranga and Torras, 2002) and analysts can talk of CS as defined from a participant perspective. Elements which are not part of the medium may have been inserted in conversation, but their presence is understood to be functional in the context in which they occurred. In other words, participants can be seen to orient to the instance of deviance as meaningful within the interactional episode.

## 2.4 What are the codes in code-switching?

The definitions of CS reviewed in section 2.2 all share one basic assumption, i.e. the fact that when engaging in bilingual conversation, bilingual speakers alternate between a discrete language A and a discrete language B. The very notion of CS as defined in grammatical studies implies the existence of two clearly-defined systems, from which speakers draw various material in interaction (Álvarez-Cáccamo, 1998: 35). Against this characterisation, it has been argued that the discreteness of languages in contact situations is a “myth” and that “code-switching should be viewed as an analyst’s construct rather than as an observable fact. It is a product of our conceptualisations about language contact and language mixing and it is not separable, either ideologically or in practice, from borrowing, interference or pidginisation” (Gardner-Chloros, 1995: 86).

The possibility of distinguishing CS from borrowing, both operationally and ideologically, is one of the tenets of Poplack and associates’ approach to bilingual speech, and has been studied in a variety of linguistic and sociolinguistic settings (Poplack and Meechan, 1998; Poplack, 2004). The proposed existence of a categorical distinction between the two phenomena clearly assumes the possibility of drawing a distinction between the two languages as they are used in conversation, and it is ultimately concerned with the language membership of single elements in speech<sup>9</sup>.

---

<sup>9</sup> It has been suggested (Muysken, 2000: 78; Stammers, 2009) that the distinction between CS and borrowing is instrumental to maintain Poplack’s (1980) general theory of CS. The classification of the overwhelming majority of lone lexical items as borrowings (whether established or nonce) effectively preserves the validity of the Free Morpheme Principle, which would otherwise be violated.

The distinction, however, is problematic to say the least. While several solutions to the problem have been put forward in the literature, there is no consensus on a boundary between the two phenomena (see Winford, 2003: 107). Several researchers have rejected the idea of a discrete boundary and placed CS and borrowing on a continuum (Myers-Scotton, 1993a; Gardner-Chloros, 2009; Treffers-Daller, 2009). They stress how the distinction should be based more on the distinction between a synchronic vs. diachronic perspective, rather than a categorical difference in synchrony.

The acknowledgement of a certain degree of “fuzziness” (Gardner-Chloros, 2009) in CS and language contact in general, however, should not be taken to mean that no boundaries or functional differences can be drawn in speech events in which speakers are using two or more languages. Rather, the challenge for researchers is to recognise that speakers, whether in relatively stable or recently formed communities or social groups, can alternate meaningfully between systems (or codes), without necessarily resorting to the notion of language as a self-contained system. In order to investigate how speakers can achieve this, it becomes necessary to temporarily abandon the view of CS typical of structural/grammatical studies and shift to more interpretive characterisations of the phenomenon, such as the conversation-analytic work reviewed in section 2.3.3.

#### **2.4.1 Languages, codes and speaker categories**

As already mentioned, researchers adopting a CA perspective take a rather sceptical stance on the “pre-existence of two or more distinct languages from which speakers draw to produce ‘code-switched’ or ‘mixed’ output” (Álvarez-Cáccamo, 1998: 35). The observation that two juxtaposed items belong to different languages is not sufficient to warrant the occurrence of CS, since speakers may not consider those items as belonging to different systems.

Similarly, in order to determine whether an alternation between two systems or ‘codes’ is taking place, speakers themselves must be shown to somehow react or align to the “otherness” of the switch (Alfonzetti, 1998). According to Heritage (1994: 128-129) it is speakers’ implicit (or explicit) understanding and

analysis of what is happening in the interaction that constitutes the main type of evidence for the analyst to claim that the alternation between two systems is meaningful. The co-occurrence of items that the external analyst sees as belonging to discrete varieties is not sufficient to argue that, from a participant perspective, CS has occurred. At the risk of over-simplifying, if no meaningful (i.e. functional) alternation between two codes can be inferred from speakers' verbal action, speakers may be using a "mixed code", where alternations have become structurally sedimented. Auer (1999) postulates a continuum between "functional" CS, language mixing and the emergence of a new mixed code, in which single instances of alternations have lost all discourse function. Language alternation for stylistic, pragmatic or rhetorical purposes is but the first step along this continuum. CS occurs when the alternation is seen as contributing to structuring the interaction or showing speakers' preferences. Attributing specific functions to single instances of language alternation, however, is not always possible, since the evidence is limited by what the analyst can reconstruct of speakers' orientation, based on a close examination of their behaviour. If the idea that the languages or 'codes' of a bilingual conversation may be unambiguously determined by the analyst is abandoned but, at the same time, it is maintained that language alternation is not a haphazard phenomenon, one needs to assume that speakers are referring to categories other than the analyst's labels "Language A" or "Language B" to interpret their own verbal actions and make sense of the on-going interaction. The analyst has to try and infer any categories based on speakers' reactions and alignment to their own speech, as can be reconstructed from close analysis of the interaction itself.

In order to uncover the categories employed by speakers in interaction and the boundaries of the codes that they employ one needs to look at the data afresh, with no prior theoretical categories in which to fit them. Instead, by carrying out a fine-grained analysis of language alternation in context, the analyst must try to uncover the very procedures used by speakers to conduct their interaction and show the ways in which their behaviour shows the mutual understanding of their utterances (Auer, 1984). The main source of evidence for the analyst is not extra-linguistic knowledge of the speakers or their sociolinguistic makeup (for instance,

any indexical value associated with the use of particular varieties or linguistic forms in a speech community) but rather is represented by the means that speakers appear to use to produce order and tackle linguistic and organisational problems that arise in interaction.

Speakers' categories, therefore, are to be regarded as something that is "brought about" by speakers themselves in conversation (see Li Wei, 1998). Categories and codes are constituted and constantly redefined in interaction, and should not be seen as something existing autonomously outside of the conversation.

#### **2.4.2 The issues of order and the base language of the conversation**

Because CA is not a unitary theory, the findings of previous work on CS do not necessarily constitute an organic body of research. Different works focus on slightly different aspects of the organisation of talk, although interest seems to be constantly revolving around the issue of 'order' in bilingual conversation (see Gafaranga, 2007). Having problematised the equivalence of codes and languages in CS, it seems logical to ask the question of what bilingual speakers refer to when communicating with one another. Order is seen as a scheme of interpretation of the interaction, a framework against which verbal actions may be seen as either adhering or deviating from a norm. In the absence of a shared interpretive framework, communication would simply break down.

Closely connected to the issues of order, otherness and deviance in bilingual speech is that of the 'base language' of the conversation, i.e. the assumption that at any given point, speakers are using only one of the languages. The difficulty in determining the base language of a conversation is discussed in Auer (2000), where several of the hypotheses put forward in the literature to determine it are reviewed. Auer argues that there is no satisfactory way of determining the base language of a conversation. Counting the number of morphemes is too basic a criterion, as is equating the base language with the language speakers are most proficient in. The notion of an "unmarked code" (Myers-Scotton, 1993b) derived from more sociolinguistic accounts of CS may be a valid criterion, but only in those contexts in which language choices are socially constrained. The very idea of a base language

of the conversation may be difficult to maintain (see also Auer, 1999).

Rather than trying to focus on whole interactional episodes, Auer claims, it may be more appropriate to look at language choices sequentially (on a turn-by-turn basis) consistently with CA methodology, and examine speakers' responses to instances of language alternation (see Auer, 1998).

Auer therefore proposes a radically localist idea of order in interaction, where the meaningfulness of language alternation relies and builds upon the context created by the language of the immediately preceding turn. Against this background, the speaker who is holding the floor has to make a choice: conform to the preceding turn in terms of language choice, deviate from it or adopt a strategy of neutrality, thus leaving a degree of ambiguity regarding the base language being used.

The idea that the interactional episode may have an overall order in the form of a base code is therefore dismissed. In order to "do justice to bilingual participants' conversational practices" (Auer, 2000: 137) the codes used by speakers to make sense of the interaction have to be seen on a turn-by-turn basis.

### 2.4.3 The medium of the conversation

Unlike Auer, Gafaranga argues that an overall scheme of interpretation in an interactional episode (or a base) needs to exist in order for conversation to take place. An overall base of the conversation is necessary for speakers to be able to interpret their own contributions as either adhering to or deviating from a norm. In the absence of a base, speakers would be unable to make sense of the on-going interaction. Evidence that the problem is relevant for speakers is found in what Auer (1998: 8) calls "language negotiation sequences", in which speakers are seen to explicitly or implicitly negotiate a "language-of-interaction" (Auer, 1998). This base is what Gafaranga calls the 'medium'<sup>10</sup> of the conversation. The medium represents the norm, or scheme, against which speakers can make sense of each other's contribution. To function as a scheme of interpretation, the medium of a conversation must be seen as orderly by participants; if that

---

<sup>10</sup> The term "medium" is used to indicate a linguistic code. The medium is only one of many codes employed by speakers in conversation (other codes including gaze, prosody, body language etc.).

were not the case, communication itself would not be possible. Gafaranga (2000; 2007) starts from the assumption that interaction between speakers, even in bilingual or multilingual contexts, is governed by norms, which speakers use to both construct their own contributions and make sense of their interlocutors' responses. These norms tend to go unnoticed as they are being oriented to, and can be discovered by the external observer most easily when they are breached. Against the backdrop of a norm (the medium), speakers are able to interpret their utterances as either complying with or deviating from what they have negotiated as the medium. Crucially, because the medium is constructed by participants, it can be made up of elements coming from more than one language.

Greer argues that

**The advantage of using the term medium instead of code or language is that it suspends the notion that same language communication is normative, at least until this can be found to be observable in the conversational data itself. Instead it is more accurate to say that, depending on the interactants, same medium communication is orderly, whether it makes use of one language or two (2008: 102).**

The discreteness of languages in CS therefore becomes blurred when examined from a conversation analytic perspective. The analyst acknowledges that conversation between bilinguals is orderly, but concedes that a framework (a base, or norm) other than the observable languages of the conversation may be used by participants. The frame need not correspond to any existing linguistic "labels". Having called into question the separation between languages, the distinction between language contact phenomena such as CS and borrowing, which is based on the possibility of unambiguously assigning language membership to single items, also becomes more difficult to maintain from a CA-based perspective. The medium of a bilingual conversation, however, should not be seen as an alternative unit of analysis to the more grammatical frameworks reviewed in **2.3.1** and subsections (many of which rely on the clause as the unit of analysis of CS), in that it belongs to a different research paradigm altogether. It presupposes a view of language which is grounded first and foremost in situated talk, rather

than in the existence of discrete linguistic systems which come into contact with one another in speech. It is therefore not directly comparable with the view of CS proposed by grammatical approaches, in that it looks at speech at a different level of analysis. The two views of language are often juxtaposed in the present work, but they should be seen as complementary, rather than mutually exclusive. While it is undeniable that much conversational and interpretive work is rather critical of the assumptions made by grammatical research on CS (see Auer, 2000; Gardner-Chloros and Edwards, 2004 for a critique of grammatical approaches), the two perspectives are both essential for our understanding of bilingual speech.

## **2.5 Code-switching involving English and Italian**

After introducing some themes in CS research which will be relevant for the analysis of flagging, the remainder of this chapter looks at English and Italian in contact situations, with a particular focus on CS. First a brief typology of contact situations involving English and Italian separately is outlined. An explanation of problems arising from the use of the label “Italian” is included. Subsequently, an overview is presented of Italian migration to the United Kingdom and of the linguistic profiles of Italian migrants during different waves of migration. Finally, previous work on English-Italian CS is reviewed. Both grammatical and more interpretive perspectives are included.

### **2.5.1 English in contact situations**

Crystal (2003: 164) claims that CS should be expected “in situations where contact with other languages is routine and socially pervasive”. Considering the undisputed position achieved by English as a global language in the second half of the twentieth century, it is hardly surprising to find English involved in a large number of contact situations including, more specifically, CS.

A discussion of the causes that have taken English to this prominent position is beyond the scope of the present work, as is an exhaustive review of studies

of CS which involve English. Crystal (2003) argues that there are two main factors, however, which should be pointed out here as driving forces behind the pervasiveness of language contact and, as a result, of CS involving English. The first factor is the history and legacy of British colonisation. With the global expansion of the British Empire, English has spread virtually all over the world, coming into contact with an enormous number of languages. The starting point of this process in the British Isles goes back several centuries, with contact between English and indigenous Celtic languages as a consequence of military and political conquest. Subsequently, this movement expanded beyond Europe into the rest of the world. America, Africa, Asia and Australia have all witnessed British presence. In some cases, this has resulted in the displacement of indigenous languages, while in others English and the indigenous languages have continued to exist side-by-side. The second factor is the rise of the United States of America as a global power and, connected to this, economic and political migration into the country in the nineteenth and twentieth century. The United States has been the recipient of tens of millions of people from Europe, Africa and Asia, all of whom have brought their linguistic repertoires into the country. The fact of suddenly finding themselves as linguistic minorities in an otherwise English-speaking environment has inevitably resulted in language contact, while the pervasiveness of this contact has – in some cases at least – created the ideal conditions for the emergence of CS. The same phenomenon has in more recent times affected other Anglophone countries: Australia, Canada and the United Kingdom have all become destinations of migration and now include numerous linguistic minorities within their borders. As a result, an unprecedented number of people have added English to their repertoires.

One peculiarity about English in CS situations is that, due to the particular sociolinguistic and historical conditions in which contact has happened, it has hardly ever been a minority language. Even in colonial settings, in which the language would be spoken by just small groups of speakers, and could be quantitatively qualified as a minority, English was the language of the coloniser and therefore in a more powerful position than the languages it came in contact with. This implies that English was (and often still is) seen as a high-prestige



language and as a desirable skill to promote personal economic advancement and social mobility – in spite of some ambivalence for its past as the language of a foreign ruler. This is for instance the case for CS in many African (Myers-Scotton, 1993a, 1993b; Eze, 1998; Finlayson *et al.*, 1998; Owens, 2005; Stell, 2009) or Asian post-colonial settings (Chan, 2003).

In the context of mass migration to the US and other Anglophone areas, English is the majority language towards which immigrant groups inevitably tend to shift in order to assimilate and improve living conditions. Different attitudes towards ethnic and linguistic minorities at different times may have accelerated the process of language shift or favoured maintenance of the heritage language(s). In Australia, for example, assimilationist attitudes and policies in the fifties have been gradually replaced by more active support and promotion of multiculturalism (Bettoni, 2010).

At some stages, most speakers would have engaged in some form of CS according to Crystal<sup>11</sup> (2003), at least in in-group communication. This is supported by the wealth of existing studies of CS as a result of migration. English-Spanish contact in the United States is perhaps the most studied language combination (see Poplack, 1980, 1988; Toribio, 2004; Smith, 2006 to name but a few), also due the size of migration from Latin America. However, a large array of language combinations has been investigated, both in the US and beyond (see for instance Clyne, 2003 for CS between English and a number of immigrant languages in Australia). Contact between English and Italian represents a similar instance of contact to those mentioned so far, even though the repertoires of Italian migrants present some peculiarities which will be discussed more in detail below.

## 2.5.2 Italian in contact situations

Compared with the global spread of English, Italian features in a much smaller number of contact situations. However, the phenomenon of mass migration out of Italy from the second half of the nineteenth century onwards has created

---

<sup>11</sup> Crystal looks at CS as a very broad phenomenon, and claims that even the borrowing of vocabulary from the language of the host country could be seen as a minimal type of CS.

the conditions for Italians to come into contact with a variety of languages and becoming, to varying extents, bilingual or multilingual. This has produced a complex range of different contact situations. In spite of a large literature existing on Italian in migration situations, CS involving Italian seems to be a relatively under-researched topic, at least in the context of Italian abroad.

There are several possible reasons for this. A first explanation lies in the fact that Italian emigration has a rather long history. This is particularly the case with the United States, which has a documented Italian presence dating back to the second half of the nineteenth century. Immigrants to the U.S.A. were under strong assimilationist pressure and may have struggled to maintain their mother tongue(s) and transmit them to generations born in the host country. With subsequent generations, the dispersion of traditional close-knit communities may have also played a part in favouring assimilation to the linguistic majority. If CS occurred in the early stages of language contact and during shift to the language of the host country, this has not been widely documented.

A different reason for the scarcity of CS research with Italian may be the fact that the focus of past work has been more ethnographic and social rather than structural or conversational. A common concern in studies on Italians abroad appears to be the problem of community language maintenance and shift (see for instance Tosi, 1984, 1991, 2001; Bettoni and Rubino, 1996; Bettoni 2010). Alternatively, the focus has been on the linguistic features of a new “ethnic” variety of Italian influenced by contact (Villata, 2003; Pietropaolo, 2010), rather than on more overt forms of language interaction, such as CS.

Finally, it should be borne in mind that interest in CS is a relatively recent development in linguistic studies. The scarcity of literature documenting CS in older communities may be due to both the impossibility of directly observing the phenomenon and the fact that other aspects of language contact – namely language maintenance or loss – attracted the interest of researchers.

### 2.5.2.1 A typology of Italian in contact situations

In the previous section the two principal factors that have brought English in contact with other languages were briefly outlined. CS situations arising from both types of contact were also mentioned. While with Italian the scale of the phenomenon is much smaller, there are still several distinct dimensions that have generated contact. Migration out of Italy is not the only factor that has caused Italian to come in contact with other languages. Other factors can be identified which have brought Italian in contact situations and have resulted in the emergence of CS.

Broadly speaking, two distinct dimensions of contact involving Italian can be identified: 1) contact outside of Italy and; 2) contact within Italy. The latter dimension can in turn be further divided in two sub-domains: contact between Italian, other Italo-Romance varieties and other historic minorities, and contact between Italian and the languages of migrants into Italy. The most well-known dimension of language contact involving Italian is intimately connected with the history of Italian migration in the nineteenth and twentieth century, and the establishment of Italian minorities across the world. It is estimated that between the second half of the nineteenth century and the 1970s over 24 million Italians left Italy to escape poverty and establish themselves abroad (Schmid, 2005). The main destinations for migrants included North and South America, Australia and Western Europe. This migratory movement has resulted in extensive contact between Italian and the languages of the host countries: English, German, French and Spanish are perhaps the most oft-cited cases.

Within Italy, as already said, there are two distinct dimensions of language contact. The first is the endogenous one, which is the result of the gradual spread and imposition of Standard Italian as a national language, starting with the Unification of Italy in 1861. Through state infrastructure, education and the media Italian first came into contact with, and gradually started competing with and partially displacing the other pre-existing Italo-Romance varieties (Tosi, 2004: 259) and the other linguistic minorities spoken within the borders of the Italian State. This has created a rather complex situation of bilingualism and language variation, with a continuum of varieties going from the local dialects to various

regional varieties of Italian (Lepschy and Lepschy, 1977), and the passage from a situation of diglossia without bilingualism to one of diglossia with bilingualism (see Fishman, 1967; Bentley, 2002). Italian, the high language that had been known by a restricted elite became increasingly used by larger sections of the population. This resulted in societal bilingualism with the maintenance of the diglossic distinction between what is perceived as a high variety (Italian) and a low one (the local other Italo-Romance variety). While the term “dialect” is often used in both the literature and lay discourse when referring to the Italo-Romance languages other than Italian spoken within the Italian State, the term does not do justice to the degree of distinctness existing between so-called “dialects” and Italian. Italo-Romance varieties are not local varieties of Italian but rather “sister languages of Italian which have developed independently from the same ancestor, Latin” (Bentley, 2002; see also Lepschy and Lepschy, 1977; Maiden, 1995). Most of these languages are not mutually intelligible with Italian and the regional varieties of the language spoken in Italy. The alternative term ‘regional language’ (Coluzzi, 2004; Tamburelli, 2010) is a more accurate descriptor of the linguistic situation of the Italian State, in which typologically similar but nevertheless distinct regional Italo-Romance varieties co-exist with Italian. Italian is an Italo-Romance variety itself but differs from its sister languages for having been adopted as the official language of the Italian State.

The second, and perhaps less well-known dimension of language contact within Italy is the one that has emerged between the end of the twentieth and the start of the twenty-first century. After over a century of emigration Italy has become a destination for both European and extra-European immigrants (Goglia, 2010). Both Italian and other Italo-Romance varieties are becoming part of the repertoire of an increasing number of non-indigenous speakers, creating new forms of multilingualism and new complex linguistic repertoires.

### **2.5.2.2 Italian varieties and Italo-Romance varieties**

When talking about “Italian” in contact with other languages, there is a risk of over-simplifying a rather complex situation. For the purpose of linguistic

description at least, the label “Italian” to describe the repertoire of migrants can be inaccurate if not misleading. For migrants leaving Italy until the 1960s the Italian language was only marginally part of the repertoire. It is reasonable to assume that, especially in the earlier stages of migration, the language was hardly known at all. While Italian already was the official language of Italy at the time, it was only known and routinely used by a very small elite. Widespread knowledge and active use of Italian was not achieved until the last quarter of the twentieth century (Maiden, 2002). If Italian featured at all in the repertoire of migrants, it only did so as a high variety, the knowledge of which was at best limited. Most individuals would have some passive competence in Italian, and be mother tongue speakers of their local Italo-Romance variety (regional language). The Italo-Romance variety would be used for all interactions within the family and in the context of the native village or town. When forced to speak Italian, most speakers would only be able to approximate it with a version strongly influenced, at the lexical, syntactic and phonological level, by the regional language, which was the speakers’ L1. This hybrid variety of Italian is sometimes referred to as *italiano popolare* (Lepschy, 2002). Upon establishing themselves in their new host country, migrants would have to quickly develop a basic communicative competence in the language of the majority. Crucially, since large groups from contiguous areas tended to migrate to the same locations, it was relatively easy to maintain the local regional language in the early stages, since migrants would reconstruct the homeland’s social network in their host country. The contact phenomena that emerged in this environment have the peculiarity of being based on a tripartite system: on the one hand the local variety(ies) of the regional language, in which speakers were proficient; and on the other Italian and the language of the majority (e.g. English or German), which were felt as prestigious languages but in which proficiency was limited. In the US, this situation gave rise to what has been called an Italo-American creole, i.e. an Italian-American version of the Italo-Romance variety with some limited influence from Italian. This new contact variety was “conspicuously affected by English lexical elements, but the main phonological, morphological and syntactic structures of the dialect remain substantially unaffected” (Tosi, 2004: 271). In other countries, such as Canada, the mixed variety has been sometimes defined as *Italiense* (Giampapa,

2001; Pietropaolo, 2010). In Canada, at least, the fact that immigration largely took place in the second half of the twentieth century meant that speakers were somewhat more competent in Italian, even though the local Italo-Romance variety remained the mother tongue of the majority.

In Australia, which has also seen mass migration in relatively recent times compared to the USA, scholars have sometimes spoken of a new variety of Italian (Australian Italian) characterised by the presence of numerous adapted lexical borrowings from English (see Bettoni and Rubino, 1996).

None of these language contact situations have been described as involving CS. Discussing the status of these contact varieties Schmid (2005: 144) speaks of “language intertwining” and argues against the definition of English/Italian contact in the three countries above as a mixed language. He contends that while there has been extensive borrowing from English the Italo-Romance grammatical structure of the bilingual speech remained virtually unaffected, thus also contesting the idea of a creole. Schmid also distinguishes these phenomena, which result from limited or imperfect acquisition of the L2, from CS. CS is seen by Schmid as a feature typical of the second generations<sup>12</sup> and more balanced proficiency in English together with Italian and/or the regional language.

With the passage from the first to the second and third generation, Italian communities in most countries have undergone the familiar patterns of improvement of socio-economic status and increased mobility. The change in conditions caused a gradual loosening of the social ties typical of first-generation communities and a gradual geographical dispersion of speakers. All these factors had an impact on language transmission: once the boundaries of the community were relaxed, opportunities for exposure to the regional language would necessarily dwindle and become relegated to interaction within the family. To these dynamics, external social pressure for assimilation should be added (see Schmid, 2005; Bettoni, 2010), as forces that contributed towards a shift to monolingualism in the language of the majority, or the maintenance of vestigial and more symbolic uses of both the Italian and Italo-Romance repertoires.

---

<sup>12</sup> It is important to note that some studies (e.g. Kinder, 1988; Bizzoni, 2003; Rocchi, 2008) have documented CS in first generation immigrants. As discussed in 2.5, the presence or absence of CS may be due at least partly to how scholars conceptualise language contact or the ways the methods used for gathering data.

### 2.5.3 Italians in the United Kingdom

Italian migration to Britain is of far more modest size compared with that to the USA, Argentina or Germany. Unlike other ethnic minorities in Britain, the Italian presence never became particularly conspicuous (Cervi, 1991). It is calculated that approximately 260,000 Italians moved to Britain between 1876 and 1976 (Schmid, 2005). Features of the migration to the UK did not differ substantially from that of other countries, except for the fact that the migration flow, while smaller, was probably more continuous, going back several centuries.

Not much is documented about the repertoires of pre-Second World War immigrants, while more is known about post-war emigration. Similarly to migrants to the New World, the linguistic repertoire of Italians emigrating until the 1970s before arrival in Britain was represented by the local Italo-Romance variety plus some limited competence in *italiano popolare*. The latter was not felt to be part of the community repertoire, and was reserved for interactions with Italian institutions or outsiders (Cervi, 1991: 217). Coming from a mostly rural background and moving to an urbanised area, these speakers frequently resorted to adapted transfers from English to describe all the new ideas and concepts that were part of their new lives (Tosi, 2001: 228). The local dialect lacked the vocabulary to describe the new reality, while the existing Italian forms were simply not known. In some cases, these adaptations were coincidentally homophonous with existing Italian words, even though they were used with different meanings from the Italian ones. English terms like ‘mortgage’ became *molleggio* (lit. ‘flexibility’, vs. Italian *mutuo*); ‘to go on strike’ *straccare* (lit. ‘to tire’, vs. Italian *fare sciopero*). Also very frequent were the cases of “false cognates”, i.e. semantic extensions of known Italian words: ‘library’ became *libreria* (lit. ‘bookshop’, vs. Italian *biblioteca*), ‘to pretend’ became *pretendere* (lit. ‘to claim’, vs. Italian *fare finta*). Communities that were geographically concentrated favoured the maintenance of the Italo-Romance part of the repertoire and sometimes created a *koiné* resulting from the contact between related Italo-Romance varieties. Families scattered over larger urban areas, on the other hand, assimilated more rapidly to English (Tosi, 2010). The influence of English on the Italian repertoire, besides the aforementioned transfers and

relexification phenomena, was mostly represented by lexical borrowings, which were morphologically and phonologically adapted. The rise of CS, on the other hand, seems to be associated more (although not exclusively) with the passage to the second generation (Gardner-Chloros, 2007; Schmid, 2005), which exhibited a more varied pattern of language interaction. This is shown, for example, in inter-generational conversations amongst London Italians (Panese, 1992).

For Italians who moved to the UK from the 1980s onwards, Tosi speaks of a new phase of Italian migration, sometimes referred to in Italian institutions as “*emigrazione tecnologica*”, i.e. ‘technological migration’. Unlike migrants from previous decades, who were mostly escaping poverty, the new migrants have considerably higher socio-economic status, and are significantly more qualified and educated than their predecessors. Increasingly this group is constituted of young professionals who are relocating abroad within the general framework of increased mobility within the European Union due to greater political and economic integration. More often than not, the new migrants relocate to the UK with at least L2 knowledge of English. Unlike their predecessors, they are native speakers of Italian and have varying proficiency in the regional language spoken in their area of origin. The largest Italian communities in Britain are located in urban areas. London has the largest Italian group, followed by Manchester and several cities in the Midlands, with smaller communities. Particularly well-known is the case of Bedford, whose Italian community has also been the subject of a book-length study (Tosi, 1984). Bedford Italians are very representative of pre-1980 migration to the UK in terms of their community dynamics and their linguistic behaviours. The first generation has the local Italo-Romance variety as its mother tongue, with limited competence in English and Italian.

The second generation, on the other hand, is dialect-dominant until school age, after which English becomes the dominant language. Italian is only made available to the second generation through formal instruction by Italian institutions or associations.



#### 2.5.4 Previous studies on English-Italian code-switching

CS studies involving English-Italian have been conducted in several of the communities seen in the previous sections, both in the UK and beyond. Consistently with the rise of interest in CS, these studies have mostly looked at communities that have established themselves between the 1950s and the 1980s, rather than older communities. The speakers therefore tend to be either first or second generation migrants.

CS in these groups has been investigated through a variety of theoretical approaches, including both structural and interpretive perspectives.

In the field as a whole, however, it seems that identity-related issues and general macro-sociolinguistic community dynamics have been the principal focus of interest, rather than the grammatical or conversational aspects of CS.

From a methodological point of view, most past research on English-Italian CS shares some common features. In several studies the data were collected through interviews with the researchers, (typically a native speaker of Italian) who would often explicitly choose Italian as the language in which to conduct the interview (that is the case in Kinder, 1988; Panese, 1992; Scaglione, 2003; Rocchi, 2008). This approach is consistent with the focus of those studies, which is mostly on the use of Italian and its maintenance, rather than language alternation *per se*.

Other researchers have looked at intra-generational semi-spontaneous interaction in the family context as participant observers (Rubino, 2003; Pasquandrea, 2007), thus somewhat limiting the influence of their presence on patterns of language use and/or interaction, at least compared with the interview setting. Yet others have investigated patterns of language use in Italian-American associations, using a mixture of interviews, group recordings and ethnographic observation (De Fina, 2007). Cases in which the researcher is completely absent from the recording session seem all but non-existent: in fact, the researcher is often the informants' interlocutor or, at best, a participant observer.

#### 2.5.4.1 **Structural/grammatical studies**

Amongst the more structural, quantitative studies, the occurrence and distribution of English discourse markers has been investigated by Scaglione (2003) in the Italian of first and second generation migrants in San Francisco, based on a corpus of semi-structured interviews. In the first generation English discourse markers used in Italian contexts are relatively frequent, but the Italian homologues have not been displaced and are also routinely employed. The second generation is marked by a significant increase in the use of English discourse markers, with some forms, such as *you know*, displacing the Italian equivalents altogether. The author argues that while most of the first generation speakers show evidence of using English discourse markers as occasional transfers into Italian, in the second generation the displacement of Italian discourse markers offers some support to the idea of a fusion between the two languages.

While it adopts a quantitative perspective, the study is perhaps not typical of grammatical approaches to CS in that it looks at discourse markers rather than CS patterns at large or specific CS constraints. Grammatical studies of CS involving Italian exist in the literature, but they involve contact with languages other than English. Amongst these, Bizzoni (2003) has applied Myers-Scotton's (1993a, 2002) Matrix Language Frame Model (see section 2.3.1.1) to naturalistic Italian-Spanish data collected among first-generation L1 Italian immigrants to Mexico. The results substantially support the predictions of the model and also provide evidence of limited attrition of the speakers' Italian. Cantone (2007) has looked at CS patterns in the context of bilingual language acquisition involving Italian-German children and argues that the predictions of the Minimalist Program (see section 2.3.1.2) are upheld in her data.

#### 2.5.4.2 **Interpretive studies**

Relatively more work has been carried out on CS involving English and Italian from a more interpretive and conversational perspective. Several studies have adopted Gumperz's idea of CS as a contextualisation cue (see 2.3.2.4), and have informed it with elements from Auer's application of CA methodology to CS,

as discussed in section **2.3.2**. While studies involving languages other than English exist in the literature (see for instance Oesch-Serra, 1998; Auer, 1995; Franceschini, 1998), focus here will be on English-Italian CS.

English-Italian CS among first and second generation Italians in London has been analysed following Auer's CA approach to bilingual conversation in a corpus of semi-spontaneous interviews and talk in the workplace (Panese, 1992). The study shows how CS in this community is used with both a discourse and a participant-oriented function and demonstrates its use as a contextualisation cue to help with the interpretation of the interaction. Speakers routinely use language alternation to signal their linguistic preference, e.g. requesting to use only one of the languages, while in other contexts they employ CS to structure the interaction, i.e. signal a change in the constellation of participants, or mark changes in footing. A similar analytical approach has been adopted in Pasquandrea (2007), in which the relationship between CS and identity is investigated. The study looks at naturalistic multi-party conversations in family settings in the United States and Australia. Three generations are normally present in the recorded data: the first generation, born and brought up in Italy and the second and the third, typically born in the host country. The repertoires of the family include an Italo-Romance dialect as well as English and Italian.

The analysis gives back a complex picture, with varying patterns between and within generations, and a series of identities which are dynamically constructed in interaction through the use of different languages. While some structural regularities can be found, the author finds no stable links between language choices and identity of the speakers, or between one language and one type of conversational activity. While speakers from all generations use language alternation to affiliate with or dis-affiliate from the various identities at their disposal, this process is achieved locally, and not by resorting to the indexical value of the languages. For example, the use of Italian does not necessarily signal affiliation with Italian identity, and, conversely the use of English does not univocally mark dissociation from Italian values.

A different study (Kinder, 1988) looks at the occurrence of discourse phenomena such as pauses, hesitations and reformulations in the Italian spoken by first-

generation immigrants to New Zealand in semi-formal interview settings with a researcher. Approximately half of the switches are produced with what the author calls “transference markers” (1988: 4), i.e. “verbal devices available to speakers for foregrounding transfers”. The study finds that the occurrence of these discourse phenomena correlates with linguistic insecurity and proficiency in English. Speakers with lower self-reported proficiency tend to produce these markers more often than speakers with higher proficiency. Kinder (1988: 8) also observes that speakers sometimes actively use these phenomena to “make a comment, more or less consciously and more or less explicitly, on a transferred item in their speech.”

In other words, according to Kinder, speakers use these discourse phenomena to monitor their own interaction. He argues that these elements also attempt to elicit the interlocutor’s feedback and response about the use of a particular item in speech and contribute to “negotiate a mutual consensus as to what is comprehensible and acceptable linguistic behaviour” (1988: 10). Kinder observes that the presence of transference markers seems to be much higher in interview settings than in the more in-group interactions by the same informants.

This suggests that the register or the formality of the situation may have had an impact on the rate of discourse phenomena produced by speakers.

Kinder does not qualify the transference markers in his data as instances of flagging, even though, at a superficial level at least, there are strong similarities between the discourse phenomena he looks at and what Poplack (1988: 218) calls flagged CS. The functions attributed to transference markers and flagging in Poplack’s work are not completely identical. The difference between the two phenomena, however, seems to be terminological rather than substantial, as will become clear in the next chapter, which focuses on the phenomenon of flagging in CS.

## 2.6 Conclusions

CS has been broadly defined in the present chapter as the use of two languages in the same speech event. In spite of former characterisations of the phenomenon as an aberrant manifestation of bilingualism, CS is now seen as a central area of investigation in linguistics and related fields. The increased interest for CS has resulted in a proliferation of theoretical approaches and definitions of the phenomenon. Some of the existing definitions of CS have been reviewed, and the similarities and differences between them compared.

Theoretical approaches to CS have also been discussed. Amongst the large body of existing literature on the subject, two strands of research have been isolated which are relevant to our work. The first strand is that of grammatical studies. The second one is that of work adopting a more interpretive stance, particularly work adopting a CA methodology. While structural studies often seek to identify grammatical constraints on CS, interpretive studies are more interested in reconstructing speakers' own understanding and use of the linguistic resources at their disposal in actual conversations.

Closely related to this theme is the issue of whether or not CS involves the juxtaposition of two separate linguistic systems or if the separation of languages in bilingual speech is no more than an external conceptualisation, with no basis in bilingual speakers' observable behaviour.

While the separation of languages is one of the tenets of grammatical studies, in more interpretive research the very possibility of identifying two discrete systems is called into question, particularly if it is not supported by the reconstruction of participants' understanding of the interaction. This is particularly, but not exclusively, the case for work conducted from a CA standpoint. Related to the possibility of distinguishing the two languages in conversation is the possibility of distinguishing CS from other contact phenomena such as lexical borrowing, and of identifying the base language of a bilingual conversation. In light of the insights from qualitative (specifically CA) work, all of these assumptions may appear quite problematic. Grammatical and interpretive work, however, adopt different views of language and language interaction, and provide complementary, rather than mutually exclusive insights into the nature of bilingual speech.

Since the present work deals with English-Italian data, an overview was given of English and Italian in CS situations. As a global language, English features in the repertoires of countless bilingual communities, and a rich literature exists investigating CS between English and other languages. One relevant feature of English in contact is that it is almost always the language of the majority and/or of the economic and political elite, unlike some of the data that will be presented in this work.

Italian, on the other hand, has a very long history as a minority language in migration contexts, both in English and non-English speaking countries. As well as contact outside Italy, Italian is found in contact situations within Italy, both with established minority languages and the new languages of migration into Italy. Work on CS involving Italian, however, is fairly limited compared with the existing literature on English. This is probably to be attributed to the orientation of scholars working on Italian, who have favoured other aspects of language contact, rather than to a genuine absence of CS. In recent times, some situations involving English-Italian contact have been described in terms of CS. Some of these have been reviewed, including both grammatical and more interpretive approaches. ¶

## THE STUDY OF FLAGGING

Searching for an exact definition of flagging in the existing literature on bilingualism, and more specifically CS, is not an easy task. The concept of flagging has received very little attention and unlike other phenomena, such as CS, little or no debate can be found in which alternative characterisations of the phenomenon are defined. This is probably due to the fact that flagging is an ancillary phenomenon which does not necessarily appear whenever CS is produced or which has not been considered relevant to grammatical studies of CS. Even where it is mentioned in the study of CS, flagging is virtually never the main theoretical focus of the research in which it features. Rather, it is most typically mentioned as co-occurring with other language contact phenomena, or as a diagnostic tool that may help distinguish different patterns of language mixing. Perhaps due to the lack of interest in the phenomenon, the definitions encountered in the existing literature are often rather approximate and idiosyncratic. As we shall see, the comparison of several existing definitions reveals that flagging is most often seen as a composite phenomenon with no exact boundaries, ranging from pauses and false starts to rather explicit forms of metalinguistic commentary. The languages involved and the features of the data may also influence a researcher's characterisation of flagging, and may lead to the inclusion or exclusions of different phenomena.

In the present chapter, several definitions of flagging are reviewed, together with an adumbration of the broader research context in which they appear.

Subsequently, a typology of flagging in CS is presented (based on Hlavac, 2006), which will constitute the basis for the classification adopted in the present work. Existing studies of CS are then reviewed which have to various extents made use of the notion of flagging. It emerges that the concept has mainly been used when distinguishing between contact phenomena, such as CS and lexical borrowing, classifying different patterns of language mixing, or describing the strategies used by bilingual speakers when dealing with potential conflict sites, i.e. sites in which the order of constituents in the two languages is not the same.

As will become evident during the discussion, however, most types of flagging are not exclusive to bilingual speech. While their characterisation as flagging devices stems directly from being produced in the environment of a switch, their use arguably derives in part from their existence in monolingual speech as well. After reviewing flagging in CS, therefore, a wider net is cast, and some of the discourse phenomena, such as hesitation and repair, that collectively constitute flagging are investigated in their occurrence outside CS contexts, in monolingual speech. The link between hesitation and repair phenomena on the one hand, and flagging on the other, is then explored in more depth.

### **3.1 Towards a definition of flagging**

One of the earliest uses of the term “flagging” in CS research is encountered in Poplack (1985; 1988). In her comparison of patterns of language mixing in two communities, she notes that instances of language alternations in her Ottawa-Hull French-English corpus a) are often accompanied by a number of discourse phenomena, such as repetitions, translations, disfluencies and hesitation markers; b) often seem to carry out a stylistic or discourse function; and c) normally show speakers’ full awareness that a change of language has indeed taken place. She claims that at least half of the switches in her French-English corpus are characterised by one or more of the following features: switches can a) provide the “apt expression” or “*mot juste*”, b) “occur while discussing languages or engaging in metalinguistic commentary”, c) bracket or call attention to the use of the other-language, d) occur while explaining, specifying or translating (Poplack,



1988). These are illustrated in examples (5) to (8). The switch to English is marked in bold typeface.

5. C'est un – a **hard boiled-killer**.

*'he's a – a hard boiled-killer.'*

6. Mais il dit, "c'est dur pour nous-autres; le, la, les, vois-tu? Eux autres, c'est rien que **the**".

*'But he says, "it's hard for us: le, la, les, you see? They only have the.'*

7. J'ai accepté le Seigneur là, ben... j'étais comme sur un... **cloud nine, cloud nine** qu'ils appellent.

*'I accepted the Lord then, well... I was like on a... cloud nine, cloud nine, as they say.'*

8. J'ai acheté une roulette, un **mobile home** là, une maison mobile.

*'I bought a trailer, a mobile home then, a mobile home.'*

(Poplack, 1988)

All of the above examples are accompanied by discourse phenomena: the switch is marked by either hesitation (5 and 7), a translation (8), explicit metalinguistic commentary (6 and 7), as well as repetitions and retracing. The presence of these devices shows, sometimes unambiguously, speaker's awareness that an other-language element has been used.

In the same study, Poplack contrasts this type of switching with that of Puerto Rican Spanish-English bilinguals in New York. The switches by the latter group of speakers are for the most part "smooth" i.e. not signalled by discourse or metalinguistic phenomena. Speakers seem "unaware" of single instances of CS, which are not seen as carrying a specific stylistic or discourse function. Rather, Poplack argues that it is the very fact of skilfully and smoothly using two languages which may globally have value as a mode of communication or a stylistic register.

This type of linguistic behaviour is labelled by Poplack as "smooth" or "true CS". That of the French-English speakers, on the other hand, is described as "flagged" CS. As already mentioned, flagged switches "are marked at the discourse level by

repetition, metalinguistic commentary, and other means of drawing attention to the switch” (Poplack, 2004: 38). Rather than a precise phenomenon, therefore, flagging is defined as a series of discourse phenomena which define a certain manner or style of CS.

In a different study (Poplack *et al.*, 1989) on Finnish-English bilingual speakers in the U.S., the concept of flagged CS is used to describe insertions of single English-origin nouns in otherwise Finnish speech. The authors mention “the presence of various discourse phenomena in the environment of the English-origin items which might indicate poor integration into host-language discourse. Included here are perceptible pauses, [...] and false starts [...] preceding or following the noun in question, and the presence of material repeating, translating or explaining the English-origin form” (1989: 393). The authors of the study also talk about the occurrence of “*flags*, forms preceding the borrowed word, which in some sense bracket or highlight it, thereby calling attention to its presence” (1989: 394, italics in the original). This introduces some degree of complexity when trying to define the concepts of flagging and flagged switch precisely. On the one hand, flagged CS is simply seen as the alternative to smooth CS, due to the presence of any kind of discourse phenomena or the interruption of the speech flow. On the other hand, there are some discourse phenomena which, arguably more than others, show speakers’ awareness of the fact that an other-language element has been inserted in speech – be it just a one-word insertion. To put it differently, according to Poplack there are discourse phenomena such as pauses, silences or hesitation markers which simply highlight a switch by their presence (i.e. make it “flagged”) and there are flags proper, which more explicitly point to or comment on an adjacent switch. Expressions like those used in example (7) or equivalents, such as *as they say*, *so-called* are part of the latter category. Not all phenomena preceding the switch may thus have the same signalling force or carry the same amount of metalinguistic information.

Turpin (1998: 225) defines flagging as “tags [...], hesitations or pauses to attract attention to the codeswitch”. She relates the presence or absence of flagging as an indicator of integration of lexical items at the level of the discourse, similarly

to what is suggested in Poplack *et al.* (1989). Unlike in that study, however, the occurrence of explicit metalinguistic commentary or translation is not mentioned as falling within the definition of the concept of flagging.

Muysken (2000) characterises the notion of flagging in a slightly different manner. He describes flagging in terms of interruption of the flow of speech at the transition point between the two languages. This interruption is prototypically achieved with a pause or the insertion of a discourse marker (such as *uhm* or *well*) preceding the switch. The result is that of effectively “separat[ing] the two languages” (2000: 101). Muysken also points out, however, that in some CS contexts, the insertion of a dummy element could equally be characterised as a flagging device. He gives examples taken from Finnish-English CS, in which a semantically vacuous Finnish element, *semmonen* ‘such’ is inserted before an English noun.

9. Siellä oli semmonen river  
there was such river

*‘there was a river’*

(Halmari, 1993: 1062; in Muysken, 2000: 106)

Muysken explicitly claims that flagging is not a unitary phenomenon. What he calls “true cases of flagging”, such as those found in Poplack’s French-English data, are different from the dummy insertions in example (9). He maintains that, rather than interrupting the flow of speech and separate languages, the insertion of a dummy element before the switch serves more as a neutralisation device, something that helps the speaker overcome linearisation problems when mixing typologically different languages. In the example above, the insertion of a dummy gives the speaker the opportunity to not assign Finnish case to the English noun *river*, therefore allowing it to occur as a bare form.

The idea that flagging may be produced to avoid sites of overt grammatical conflict is also found in Gardner-Chloros and Edwards (2004: 108), where it is stated that flagging involves “inserting a conversational marker or comment at the point where the switch occurs.” These devices are seen as ways of side-stepping or neutralising any “grammatical awkwardness resulting from switching at a particular point in the sentence”. Interestingly, attention is also drawn by the

authors to the fact that many of these strategies, amongst which they include repetitions, interruptions and reformulations, are not exclusive to CS, but are also routinely used functionally by monolingual speakers, and help structure the message to be conveyed. They should also be seen as functional when appearing in a bilingual environment. In bilingual speech, besides, they argue that these elements can also contribute to legitimise combination between languages which would otherwise not occur.

While he does not give an explicit definition of flagging, Hlavac (2006) observes in his work on different generations of Croatian-English bilinguals in Australia how some discourse markers, such as English *like* or its Croatian equivalent *kao*,<sup>13</sup> can be used “as a signal that ‘other language text’ may follow” (2006: 1893). He describes this signalling use of discourse markers as “forewarning” the interlocutor. This strategy seems to be adopted especially by second-generation speakers interacting with first generation speakers in his data. The former often employ flagging devices to alert their interlocutor that a switch is about to take place. In this sense, Hlavac’s view of flagging is similar to Poplack’s. Both characterise flagging as elements that draw the attention of the listener to the switched word or segment.

### 3.1.1 A typology of flagging

Hlavac is possibly the first to draw up an explicit typology of flagging strategies, based on his Croatian-English data. The list of phenomena considered to be flagging is reproduced below. The original list (Hlavac, 2006: 1894) makes explicit reference to Croatian and English as the languages of some of the elements – as in the case of lexicalised filled pauses (A.iii). These have been replaced here with “same language as switch” and “other language”, which normally means the language used on either side of the switch.

---

<sup>13</sup> Hlavac (2006: 1895) notes that this use of *kao* is not “widely attested in homeland Croatian.” It therefore seems to be a feature of the variety of Croatian spoken by Croatian-English bilinguals in Australia.

### **A. Phenomena preceding switch**

- i. Paralinguistic markers (e.g. laughter, nervous coughing)
- ii. Unfilled pauses
- iii. Lexicalised filled pauses (*like, you know*)
  - a. Same language as switch
  - b. Other language
- iv. Non-lexicalised filled pauses (*uhm*)
- v. Other-language equivalents
  - a. Exact synonyms
  - b. Approximate correspondents
- vi. Explicit pre-empting
  - a. Same language as switch
  - b. Other language
- vii. Disfluencies, false starts, repairs

### **B. Phenomena following switch**

- i. Paralinguistic markers (e.g. laughter, nervous coughing)
- ii. Unfilled pauses
- iii. Lexicalised filled pauses (*like, you know*)
  - a. Same language as switch
  - b. Other language
- iv. Non-lexicalised filled pauses (*uhm*)
- v. Other-language equivalents
  - a. Exact synonyms
  - b. Approximate correspondents
- vi. Explicit justification
  - a. Same language as switch
  - b. Other language
- vii. Disfluencies, false starts, repairs

The first noticeable feature of this typology is the distinction between flagging occurring before the switch and after the switch. This sets apart Hlavac's characterisation of flagging from several of the others encountered earlier, which are based on the assumption that flagging typically signals the occurrence of a switch *before* the switch itself has been produced. The observation that flagging can occur after the switched segment as well, however, is somewhat at odds with

Hlavac's own idea that flagging has the function of forewarning the interlocutor. It may also be problematic for the idea that the function of flagging is to neutralise or avoid any overt linearisation problem when mixing typologically different languages, as argued in Muysken (2000) and Gardner-Chloros and Edwards (2004). If the function of flagging were simply that of breaking the speech flow, so that the speakers can insert elements in a clause or sentence and circumvent a grammatical constraint, it is not clear how flagging after the switch itself could fulfil this role.

Hlavac's Croatian-English data show that flagging before the switch is the prevalent pattern, with 479 flags occurring before the switch vs. 285 occurring after the switch. It is not possible to tell from his analysis, however, whether a) the same switched segment can be flagged by more than one of these elements placed one after the other (i.e. FLAG FLAG SWITCH: an *ubm* followed by a silent pause or another marker) or b) whether a switch can be flagged both before and after being produced, i.e. FLAG SWITCH FLAG.

While there is a difference in the quantitative distribution of flags across different categories, the types of flagging strategies appearing in Hlavac's data are virtually the same in both positions, with the partial exception of (A.vi) and (B.vi), labelled *explicit pre-empting* and *explicit justification* respectively. The former necessarily occurs before the switch, while the latter is only found after the switch. The two devices typically fulfil the same function of accounting for the other-language segment, for example through a metalinguistic commentary. Pre-empting, however, may also be seen as a way of avoiding switching altogether.

By looking at the list of flagging devices above, it is apparent that very heterogeneous elements are included under the umbrella term of flagging. These range from paralinguistic markers, such as laughter or nervous coughing, to explicit metalinguistic commentary, such as translation or justification of the switch, to phenomena which could be attributed to either the nature of speech production or lexical retrieval difficulties, such as hesitation markers, false start, and unfilled pauses.

More importantly, however, it should be noted that the typology is heterogeneous, in that it defines some elements in merely structural terms

(e.g. filled pauses), and others according to the broad function that they fulfil in speech (e.g. justification). Some types of flagging can be very easily defined structurally – typically a language only has a handful of elements that can be used as filled pauses – while others can take a wide range of forms and are not always entirely predictable *a priori* or definable other than in functional terms. That is, for instance, the case of post-switch justification.

It could be questioned at this point whether all these phenomena should at all be considered as different manifestations of an alleged unitary concept of flagging. The issue is further complicated by the fact that, with the possible exclusion of overt translations, all of the other phenomena can also be found in the speech of monolinguals (see Backus, 1996: 113) and in the speech of bilinguals when they are not engaging in CS. While Hlavac does not address this question explicitly, he notes how in his data flags “are twice as likely to occur in the vicinity of a transfer or code-switch as they are in general discourse throughout the sample” and that “there is a very large number of transfers which have flagging phenomena preceding them” (2006: 1849). If we are to treat the different manifestations as part of a unitary phenomenon, then adjacency to the switched element or segment and the interruption of the speech flow in some form seem to be the minimal shared features underlying all types of flags.

### **3.2 Previous work on flagging in code-switching**

Flagging has not been extensively analysed in work on CS, and only features in a small number of studies. Different characterisations of the phenomenon, however, are found in the literature, as shown in the discussion of various definitions in the previous part of this chapter.

In what follows existing analyses of flagging in CS are discussed.

### 3.2.1 Flagging as indicative of different code-switching styles

The occurrence of flagging in CS has been argued to distinguish between different language-mixing styles within an immigrant community and investigated to ascertain whether CS can be said to constitute “a mode of discourse in the community” or is used “meta-linguistically” (Eppler, 1994: 75). In her sociolinguistic and linguistic profiling of the Austrian immigrant community in London, Eppler (1994) maintains Poplack’s distinction between “smooth” and “flagged” as basic patterns of CS. Eppler argues that for speakers “in ‘flagged’ code-switching situations, one language is usually endowed with affective import but is regarded as having less instrumental value than the other” (1994: 83). Speakers’ attitudes to their languages may therefore play a role in determining the CS patterns found in a given population. Even though she does not specify which is which in her data, it is reasonable to assume that (Viennese) German, the speakers’ L1, is the language with the higher affective value in this group, while English is the one with higher instrumental value.

Both smooth and flagged CS are found in Eppler’s data and they are seen as indexical of two subgroups in the Austrian émigré population. One “core” group, with socialisation patterns involving frequent contacts with fellow Austrians and geographical concentration in a localised area, engages frequently in smooth CS. The other group, on the other hand, whose members are scattered geographically and are generally more integrated into British mainstream society, exhibits a preference for flagged switches. The core group, which presents the full features of a rather close-knit linguistic community, is perhaps comparable to Poplack’s (1988) Spanish-English New York speakers, for whom CS constitutes an established mode of communication and part of their linguistic repertoire.

The speakers in the non-core group, on the other hand, are more similar to French-English speakers in Canada, and they tend not to have strong ties with fellow Austrian immigrants.

Eppler’s analysis of flagged CS concentrates on the switches themselves and the discourse functions that they achieve in discourse, rather than the flags by which those functions are indicated or the switches highlighted. The classification of her data is therefore quite interpretive, with examples of switches being classified



as flagged depending on whether the discourse functions they fulfil in speech can be identified. For example, she argues that speakers using flagged CS tend to switch languages in conjunction with *mot juste* expressions, in the context of translating, explaining or providing metalinguistic commentary much more often than speakers in the core group. An instance of CS for the *mot juste* is reproduced in (10). The focus in the classification is therefore on the functional value of the switch (i.e. providing the apt expression), rather than on the presence of any overt markers signalling that a switch is about to occur or has just occurred. In fact, were it not for the presence of the discourse marker *you know* – a lexicalised filled pause in Hlavac’s (2006) typology – after the switch in (10), it could be argued that the switch is smooth, rather than flagged.

10. You know, in those days in Vienna – it probably still is a bit like that now – in order to be accepted, and be somebody, it was good if you looked like **ein Herr Direktor**,  
a Mister director  
 you know. If you went into a restaurant and looked like that, you got the best seat...  
(Eppler, 1994: 87)

To the types and functions of flagged switches mentioned above, which are essentially derived from Poplack’s work, Eppler also adds the presence of flagging devices such as pauses, hesitations and false starts (1994: 85). This group of devices is not necessarily seen as having an explicit rhetorical or discourse function, but it also contributes to marking a distinction between two different patterns of CS in the population under study.

### 3.2.2 Flagging as an indicator of discourse-level integration

Besides being indexical of different CS styles and being related to the structure of speakers’ social networks, the occurrence of flagging has been related to integration (or lack thereof) of switches into the host language, and more specifically, to the possibility of distinguishing switches from lexical borrowings. The presence or absence of flagging with single English-origin nouns occurring in an Acadian French corpus is briefly analysed in Turpin (1998). Flagging is

used alongside morphological and syntactic measurements to draw a categorical distinction between code-switched forms and borrowed forms. Turpin adopts Poplack's comparative method for distinguishing between CS and borrowing. The underlying assumption of the comparative approach is that borrowings are more integrated at various levels into the host language than clear-cut instances of CS; while the two may be operationally difficult to distinguish, Poplack (2004) maintains that they are clearly different contact phenomena. Turpin compares lone English-origin nouns in French discourse with both unambiguous code-switched multiword sequences and single host-language nouns occurring in the monolingual speech of the same speakers for both languages involved (French and English in her case). Items exhibiting the same behaviour as their homologues in the recipient languages with regards to the various diagnostics are classified as borrowings. If they pattern differently from the same items occurring in unmixed environments, on the other hand, they are classified as switches. Turpin describes the presence or absence of flagging as indicating "integration at the discourse level", to distinguish it from other levels of integration (syntactic and morphological), which include the presence of recipient language agreement markers or determiners. Turpin finds that over a third (35%) of unambiguous switches present some flagging, against only 9% of lone English nouns in French contexts. This figure is in turn compared with English nouns in monolingual English contexts (7% flagging) and French nouns in monolingual French contexts (3%). The similarity of percentages of flagging (9% and 7%) leads the author to conclude that "at the discourse level, the lone English-origin nouns are treated as legitimate members of the lexicon of the language in which they are embedded" (Turpin, 1998: 225) or, in other words, are borrowings. The presence of flags with lone English nouns in French is therefore dismissed by Turpin as a discourse marker with no value specific to language alternation, since their distribution mirrors that found for nouns in monolingual contexts.

One of the strengths of the comparative approach is that the ambiguous data are compared with alleged unambiguous data from the same speakers; in this way, it is possible to see whether other-language elements are at various degrees integrated in the variety of French spoken by those very speakers, rather than some abstract standard variety which may not be part of the speakers' repertoire.

The rate of flagging of switched lone nouns, for instance, is compared with the flagging (or presence of discourse markers) for lone nouns in different unmixed environments by the same speakers. Acquiring adequate knowledge of the speakers' particular variety (or varieties) of the language, however, may be methodologically challenging, especially in recent or unstable contact situations; while certainly a desideratum, therefore, it may not always represent a viable option.

There are at least two problems, however, with Turpin's analysis. Firstly, she denies that flagging plays any special signalling role for English-origin nouns in French discourse because their rate of flagging (9% or 55/604) is similar to that of French nouns in unmixed contexts (3% or 19/567). The analysis, however, is simply descriptive, i.e. not backed up by statistical analysis. Turpin describes the percentages as similar, but does not specify what kind of difference would be needed to claim that a difference exists.

A way of doing this, which Turpin does not mention, would be to carry out a statistical test. A simple chi-square test on her results would have revealed that the difference between these two groups is in fact statistically significant, and quite strongly so<sup>14</sup>. This highlights that there is a difference in the incidence of flagging for these two categories. Additionally, the flagging rates for lone English nouns in French are very similar to those for English nouns in unmixed environments (7% or 26/372). A chi-square test reveals that there is no significant difference between these two categories. This entails that English nouns in French discourse actually pattern more like English nouns in unmixed English than French nouns in unmixed French; this in turn indicates that at least some of them are not well-integrated at the discourse level into their recipient language. The core of Turpin's argument, however, lies in the difference in rates of flagging between lone English nouns in French and "unambiguous codeswitches", i.e. multiword English sequences which are not side-by-side borrowings or fixed expressions. These are flagged 35% of the time or 71/204. An example of a flagged unambiguous switch from Turpin's data is given in (11).

---

<sup>14</sup>  $\chi^2 = 13.5535$ ,  $df = 1$ ,  $p < 0.001$ . The test is not used in Turpin's analysis.

11. Il aime pas l'autorité – l'autorité – **you know, everything is a conspiracy**, tu sais là?

*'He doesn't like authority – authority – you know, everything is a conspiracy, you know?'*

(Turpin, 1998: 225)

The switch *everything is a conspiracy* is classed as flagged, because it is bracketed by discourse markers *you know* and *tu sais là* 'you know', which show speaker's awareness that another language is being used. Arguably, the whole sentence *everything is a conspiracy* is flagged, and not just the noun *conspiracy*. It is not clear from Turpin's analysis, however, how exactly unambiguous switches are classified as being flagged or unflagged. The example in (12) illustrates this point more clearly:

12. C'est une maison de douze personnes, pis- bien c'est j -- || **just a small farm and two people run it, like**, tu sais?

*'It's a twelve person house and well- it's – just a small farm and two people run it like, you know.'*

(Turpin, 1998: 224)

Turpin cites *farm* and *people* in the above example as instances of nouns occurring in unambiguous CS sequences. The two nouns are produced with no flagging, but the switch point at the start of the sequence is marked by an interruption of the speech flow and a pause, which could be interpreted as flagging. It can be argued that the switch point is indeed flagged. Whether single elements within the sequence should also be classified as flagged, however, is more difficult to maintain. If Turpin is counting nouns such as those in (12) as flagged, then it comes as no surprise that such a higher percentage (35%) of nouns in unambiguous switched contexts are flagged. The exact criteria for classifying nouns in unambiguous switches as flagged, however, are not explained clearly in her work. If Turpin's comparison is intended to be between elements of the same class in different contexts, then only nouns in unambiguous switched contexts with flags in their immediate vicinity should be counted as flagged, similarly to what is done with the three categories of lone nouns. If that is not the case, Turpin is contrasting single nouns against entire phrases or clauses, which arguably does not constitute a balanced comparison.

### 3.2.3 Flagging as indicative of different code-switching strategies

Muysken's (2000) typological approach to CS (see 2.3.1.3) identifies different mechanisms underlying language-mixing in bilinguals' speech. Muysken also points out how different types of discourse phenomena may be indicative of different mixing strategies operating in speech.

Turpin's (1998) "unambiguous CS sequences" have several traits in common with Muysken's definition of alternational code-mixing. As mentioned earlier, flagging plays a part in this type of language mixing in that it makes the languages "remain relatively separate" (Muysken, 2000: 96). In Muysken's system the separation of the two languages is achieved via the insertion of a pause, hesitation or discourse marker, which interrupts the flow of speech. These instances of flagging are seen as a separate phenomenon from "dummy word insertion", seen in example (9) above. He argues that the latter are inserted to solve a linearisation problem in speech production, e.g. a difference in head-modifier order in nominal constructions. True flagging, on the other hand, may be associated with the speaker's reluctance to mix languages (2000: 106), for instance as a result of external social or situational pressure to speak monolingually. Dummy word insertions are regarded by most other scholars as being part of the same phenomenon as Muysken's true flagging (see for instance Poplack, 1998; Turpin, 1998; Hlavac, 2006).

Flagging (in Muysken's sense) and dummy word insertions are employed together with several other linguistic features as diagnostics to distinguish between different types of CS (or code-mixing, as he labels it). True flagging is considered to be indicative of alternation, while the insertion of dummy words seems to indicate the insertional pattern. Muysken's diagnostic system is applied in Deuchar *et al.* (2007) to Welsh-English, Mandarin-Tsou and Taiwanese-Mandarin data. The analysis reveals that each language pair shows the preference (in quantitative terms) for one CS pattern and the presence of a secondary pattern (e.g. insertion as the prevalent pattern with congruent lexicalisation as a secondary one). The import of flagging and dummy insertions is not discussed in great detail, as these represent only two amongst a rather large number of diagnostic tools for the identification of different code-mixing strategies. It is worth noting however, that they are associated to different types of language mixing.

### 3.2.4 Flagging as indicator of a borrowing/code-switching distinction?

The occurrence of flagging has been related to discourse level-integration and as a diagnostic tool to distinguish between different outcomes of language contact (Turpin, 1998). Differences in types of flagging, however, may also be indicative of a distinction between different phenomena. This has been hinted at in section **3.2.3** with reference to different types of CS, but has also been applied by others to distinguish between CS and borrowing.

In her study of Jersey Norman French-English bilingualism, Jones (2005) focuses on the occurrence of single word English insertions in Norman French speech. Flagging is defined essentially as a device to draw attention to a switch or show awareness that a switch has been produced, following Poplack's definition. Jones investigates the occurrence of flagging with single word English insertions in Jersey French as a possible way to distinguish between CS and lexical borrowing. She finds that of the 443 singly occurring English contact forms in her data, regardless of syntactic or morphological integration, only 102 (or 23%) are produced with any flagging at all. This low figure is attributed to the length of contact between the two languages, and the high level of penetration of English into Jersey Norman French.

Flags in the data are divided into three macro-categories: hesitation, metalinguistic commentary and self-correction. Metalinguistic commentary is the most frequently used flagging strategy (43 tokens), followed by self-correction (31 tokens) and hesitation (28 tokens). Jones uses a frequency criterion to separate all of the English forms in her data into switches and borrowings and investigates whether a difference exists in the frequency and type of flagging. The division in two subcategories is based on an arbitrary frequency criterion; all forms with three or more occurrences are classed as borrowings, while those with less than three occurrences in the corpus are classified as switches. The results of this classification are presented in **Table 1**. It first emerges that a higher percentage of switches is flagged compared with borrowings (34% and 19% respectively). Further to this, half of the flagged switches (52% or 31/60) are marked by self-correction, such as the example (**13**) below, in which the English insertion *playground* is post-flagged by its translation equivalent in Jersey French *bel de l'école*.

13. Tout l'monde pâlait l'Jèrriais dans l'playground, dans l'bel de l'école.

*'Everyone spoke Jèrriais in the playground.'*

(Jones, 2005: 13)

Borrowings, on the other hand, are never flagged with self-correction, and when they are flagged they are only marked by hesitation or metalinguistic commentary.

<b>Flagging strategies used with forms with three or more uses (borrowings)</b>	
No flag	224 (81%)
Flagged forms	42 (19%)
Pause/hesitation	18 (43% of flagged forms)
Metalinguistic commentary	24 (57% of flagged forms)
Self-correction	0 (0% of flagged forms)
<b>TOTAL</b>	<b>266 (100%)</b>
<b>Flagging strategies used with forms with less than three uses (switches)</b>	
No flag	117 (64%)
Flagged forms	60 (34%)
Pause/hesitation	10 (17% of flagged forms)
Metalinguistic commentary	19 (32% of flagged forms)
Self-correction	32 (51% of flagged forms)
<b>TOTAL</b>	<b>177 (100%)</b>

**Table 1** Flagging of switches and borrowings in Jersey French (Jones, 2005)

Jones concludes that the latter result “reveals a difference in the speaker’s mental perception of the two forms and also seems to corroborate the idea that borrowed forms are more fully integrated into the recipient language than code-switched forms” (2005: 14). The results also show that, at least for the Jersey French-English data, the amount of flagging for a given inserted item diminishes with increased frequency. There is also a stark difference in the types of flagging strategies, such that “the incidence of flagging by self-correction [could function] as a possible additional criterion [besides morphological and syntactic integration] for distinguishing between singly-occurring code-switched forms and borrowings” (2005: 19).

It is worth stressing again that the distinction between switches and borrowings is marked by a completely arbitrary frequency criterion in this study, and may

therefore be questionable; what is important, however, is how in absolute terms the percentage of flagging is higher for switches, i.e. items with a lower frequency in the data.

### 3.2.5 Flagging from a Conversation Analysis perspective

The studies mentioned so far have in common the fact of having a predominantly grammatical perspective when looking at flagging. The role of flagging is framed in relation to either the identification of different CS patterns in the data or the distinction between CS and borrowing in ambiguous cases, i.e. with lone lexical items. Flagging is used as a diagnostic tool, however, also in work with a more interpretive and interactional focus, such as studies carried out within the framework of Conversation Analysis (CA).

The presence of flagging in the environment of a switch is used in Gafaranga (2000) to distinguish between different types of searches for the *mot juste* by multilingual Rwandese speakers living in Belgium and Catalan-Spanish bilinguals in Barcelona. While the author does not explicitly mention the concept of flagging as a diagnostic to identify different strategies adopted by speakers in his data, he extensively refers to the occurrence of phenomena such as pauses, hesitations and repetitions as evidence that speakers have been going through a lexical search process before a switch. These “aspects of talk which indicate that current speaker is going through a search process” are referred to as “trouble markers” (2000: 333). Gafaranga therefore corroborates the view that flagging somehow draws attention to the switch (although he would not use this terminology).

In his data, Gafaranga looks at the presence of flagging devices (or “trouble markers”) co-occurring with a switch, which constitutes evidence that the speaker is going through a lexical retrieval or formulation problem. Gafaranga also notes how on some occasions, after the lexical search has ended with the insertion of an other-language item, the speaker engages in further repair work, for instance by providing a translation of the inserted item or a justification of why it was uttered – both types of flagging in Hlavac’s typology. Gafaranga



interprets the presence of these repair sequences after the lexical search as evidence that the speaker is aware that an element that is deviant or not appropriate for the particular interaction has been inserted.

An example of this is provided in (14) below. The example is made more complex by the fact that both speakers are using three languages: Kinyarwanda (shown in plain script) French (in italics) and English (in bold). Speaker A encounters a problem, as marked by his pause at the end of turn 1. He then inserts the *mot juste* he is looking for, *local government/local authority* in turn 3. The insertion is in English and flagged, as it is preceded by the explicit pre-emptive *babyita* ‘they call it’. Immediately after the English insertion, A produces more flagging by attempting a repair of what he has just said. This is also marked by the lengthening of *ni nkaaa* ‘it’s likeee’<sup>15</sup> at the end of turn 3. The hesitation in turn triggers the intervention from speaker B, who offers a translation equivalent *municipalité* (a further flag) in turn 4. This is finally ratified by speaker A in turn 5.

14. *Participants are talking about the possibilities for a refugee to study in the UK.*

1. A: noneho rero nka bariya b’ impunzi ukuntu bigenda (.) babagira ba (.) a a amashuri hano ni *privé quoi* (.) ni *privé* mbega (.) kuburyo rero kugirango aze muri iyi *université* agomba kwishyura (.) *mais comme* nta mafaranga afite ay yatse *bourse le*

2. B: umh

3. A: babyita **local government** (.) **local authority** *donc* ni nkaaa

4. B: ni nka *municipalité*

5. A: ni nka *municipalité c’est ça* (.) *municipalité* yahano niyo yamuhaye *bourse*

---

1. A: refugees like him are (.) schools here are *private* (.) they are *private* so that he must pay to study at this *university* (.) *but as* he doesn’t have money he has had to apply for a *grant* from *the*

2. B: umh

3. A: they call it **local government** (.) **local authority** well it’s likeee

4. B: it’s like a *municipality*

5. A: that’s right it’s like a *municipality* (.) he got a *grant* from the local *municipality*

(Gafaranga, 2000: 338)

---

<sup>15</sup> In his data, Gafaranga marks lengthening by repeating the vowel in the orthographic spelling.

Gafaranga's notion of repair in this context shares many traits with Jones' idea of self-correction seen in section 3.2.4. When talking about flagged CS in her data, Jones observes that speakers may use self-correction because they are "more conscious of using these forms" (2005: 20) and may therefore perceive them as violating a conversational norm. Gafaranga also highlights how sometimes the presence of trouble markers (or flagging) may lead the speaker's interlocutor to intervene and collaborate to end the search for the *mot juste*, as it is indeed the case with speaker B in example (14) above. The repair is initiated by speaker A immediately after the switch *local government* (.) *local authority*, but it is completed by speaker B, who provides the word *municipalité* 'municipality'. Gafaranga does not specifically elaborate on the importance of this, but the intervention of the interlocutor constitutes evidence that – at least in these instances – the function and/or meaning of the flags in the immediate environment of the switch are clear to the speakers themselves, who collaborate to solve the problem that has arisen. Speaker B correctly interprets the value of speaker A's flags as a lexical search and collaborates by offering a possible solution, which is in turn ratified by A.

### 3.3 Does flagging exist in monolingual speech outside code-switching?

As has already become apparent, a number of very different phenomena typical of spontaneous speech production fall within the definition of flagging. Looking at the typology of flagging in section 3.1.1, however, it is very easy to see that, with the exception of other-language equivalents<sup>16</sup> (v), virtually all the devices that are classified as flagging are not exclusive to CS. On the contrary, they are also found in the speech of monolinguals (see Gardner-Chloros and Edwards, 2004) and, as would be expected, in the speech of bilinguals when they are not engaging in CS (see Fehringer and Fry, 2007). This may potentially pose a problem for the assumption that the phenomena referred to as "flagging" have a

---

<sup>16</sup> In fact, even this strategy may be compared to providing a synonym from a different language register in a monolingual environment. It has often been argued that style-shifting and CS are very similar phenomena, and driven by the same motivations (see Gumperz, 1982).

special function – signalling or otherwise – when occurring in the environment of a switch and are not just ordinary discourse phenomena.

In most of the CS studies mentioned in section **3.2** and subsections, flagging is treated as a loosely-defined heterogeneous phenomenon. It is sometimes conceded that there may be meaningful subcategories within it (as in the case of Jones, 2005, in which hesitations, metalinguistic markers and self-corrections are treated as partially distinct manifestations) and that not all the devices used for flagging a switch have exactly the same function. The features or functions of single types of flagging, however, are not given much attention in the CS literature. In order to gain a better appreciation of the various phenomena that collectively constitute flagging, it is necessary to temporarily abandon CS and look at the occurrence and function of the discourse phenomena that fall within the definition of ‘flagging’, but when they occur in monolingual contexts. Most of these discourse phenomena tend to be virtually the same in form in monolingual and bilingual speech; their functions, while not identical, could also be related. It is thus possible that a link exists between flagging and discourse phenomena in monolingual speech. The use of these phenomena in monolingual speech could represent the “monolingual homologue” of flagging in CS. The majority of these discourse devices are equally available to monolingual and bilingual speakers alike and are routinely used by monolinguals; it is when they are used by bilingual speakers in conjunction with a switch, however, that they can assume a more specific function.

This assumption is not to deny that flagging has an element of specificity to bilingual speech and CS in particular; nonetheless, it represents an acknowledgement that speakers do not always create new strategies *ex novo* when engaging in CS, but rather use the same resources and methods and apply them in a different domain. This assumption is implicitly present in some work on flagging, in which discourse markers in monolingual speech are compared with flagging in bilingual contexts (Turpin, 1998), but the link between the two domains has not been spelt out explicitly.

Within the field of research on discourse phenomena in monolingual speech, the areas of hesitation and repair seem particularly suited to establishing a link

between monolingual speech on the one hand, and CS on the other. While the fit is not perfect, hesitation and repair can be considered the monolingual counterparts of flagging. The similarities between repair and hesitation on the one hand and flagging on the other are visible at different levels; superficially all phenomena are quite heterogeneous and can be produced in speech through a number of devices or strategies, which tend to be for the most part the same in monolingual and bilingual speech. Also, many of these phenomena have often been treated as peripheral, perhaps because of their exclusive relevance to spoken, rather than written language. They are not considered part of the message the speaker intends to convey, but at best mechanisms related to the very nature of speech planning and production (although there are exceptions to this idea, as argued in Clark and Fox Tree, 2002).

Regardless of their interpretation, however, most studies seem to converge at least on acknowledging some degree of signalling value for most of these devices.

Some overlap can be also seen between the concepts of hesitation and repair. Crucially, flagging seems to possess the features of both phenomena.

A basic understanding of hesitation and repair in monolingual environments is therefore necessary in order to establish the link between different phenomena.

Talking about discourse phenomena in monolingual contexts, Gardner-Chloros *et al.* (2000: 1338) note how

[r]esearchers working on monolingual discourse came relatively late to an understanding of the discourse functions of forms that occur only in speech: it is not so long ago, for example, that ‘you know’ and ‘I mean’ were considered to be mere fillers and fumbles, terms reflecting the view that speakers used them only to keep the floor while planning and producing their utterance.

The bias against the functional aspects of discourse phenomena in earlier linguistic and psycholinguistic work can still be observed in the widespread use of labels such as “disfluencies”, “filled pauses” or “disruption”. These characterisations seem to underlie the concept of an “ideal delivery” of speech, a concept which finds little or no grounding in the reality of spoken language,

as a quick look at a transcribed sample of conversation easily shows (see also O'Connell and Kowal, 2005, for a discussion).

The basic value of hesitation and repair phenomena as “fumbles”, symptoms of a speech production problem, or ways of keeping the conversational floor, does not need to be eliminated altogether. It is now assumed, however, that these phenomena can also achieve other communicative functions at the same time (Gardner-Chloros *et al.*, 2000: 1338). The functional nature of most of these phenomena in the monolingual literature is now undisputed.

Unfortunately, the field of discourse phenomena is characterised by great terminological ambiguity; the fact that it has been the object of study by scholars from areas as different as pragmatics and psycholinguistics has perhaps exacerbated this problem, with the result that several only partially overlapping nomenclatures are now found in the literature. While discussing these phenomena, an effort has been made here to keep terminological ambiguities to a minimum and adopt labels that are as consistent as possible with those used for flagging in CS.

### 3.3.1 **Hesitation and filled pauses in monolingual speech**

Amongst the phenomena typical of spontaneous speech, hesitation probably has the longest-established research, going back to at least Maclay and Osgood's (1959) seminal study on hesitation phenomena in English. The authors' definition of hesitation includes several types of phenomena, such as retracing, unfilled and filled pauses. One of their main claims is the idea that hesitation phenomena in speech are not random, but patterned. For instance, most types of hesitations tend to occur before content rather than function words. Hesitations are linked to greater lexical choice and points of “highest uncertainty, points where choices are most difficult and complicated” (Maclay and Osgood, 1959: 41). In a bilingual context, the environment preceding the switch could be a point of uncertainty, since at least in some cases, speakers are faced with the choice of whether or not to use a different language from the one hitherto employed. Switches also tend to be content rather than function words – single nouns are the most oft-inserted grammatical category in a large number of CS data sets.

Within hesitation, non-lexicalised filled pauses (also referred to as “fillers”) are probably the single category of flagging phenomena which has been more thoroughly investigated in the literature (see Clark and Fox Tree, 2002 for a brief review; see also Kjellmer, 2003; O’Connell and Kowal, 2005; de Leeuw, 2007; Fehringer and Fry, 2007; Schegloff, 2010) and has been at the heart of a lively debate. Prototypical exemplars of fillers are English *uh* and *uhm* (also spelled as *er* and *erm* in British English). The nature of fillers is at the core of what has been referred to in the literature as the “signal vs. symptom” debate (de Leeuw, 2007). For proponents of the symptom interpretation, filled pauses are a direct consequence of a cognitive process on the part of the speaker, such as a speech planning issue that is being addressed – for instance a lexical search. For supporters of the signal interpretation, on the other hand, the presence of fillers is geared mainly towards the listener, and is used to communicate various messages relating to the speaker’s own performance – e.g. that a word that is potentially not appropriate or that requires special interpretation is about to be uttered (see Corley and Stewart, 2008).

The strongest case for the interpretation of fillers as signals is made by Clark and Fox Tree (2002). The authors classify *uh* and *uhm* as interjections, whose main function is to signal to the interlocutor that a further delay (e.g. a pause) in speech delivery is coming. Evidence for this claim comes from the frequent occurrence of either silence or more fillers after *uh* or *uhm* in the authors’ data. Fillers are therefore planned elements and part of the message the speaker wants to convey. Rather than being part of the main message, however, their function is to comment on the speaker’s own performance. Clark and Fox Tree also argue for a division of work for the two forms, with *uh* signalling minor delays and *uhm* marking a major delay in ensuing speech.

This interpretation is challenged in O’Connell and Kowal (2005), who, based on an acoustic analysis on a different data set, argue that the claim that fillers signal further delay is not warranted. In the majority of cases, *uh* and *uhm* are followed by no pause at all in their data. In those cases in which a pause occurs, its duration does not differ depending on which filler was used, leading them to reject the claim that *uh* and *uhm* signal two different types of delay.

The authors concede that *uh* and *uhm* should be treated as genuine words, although their data and analysis do not support a “delay-signalling” interpretation and the morphological classification of fillers as interjections. They do argue however that fillers “far from disrupting and suspending speech, protect and sustain genuine fluency” (2005: 572), or in other words, have a communicative value. Corley and Stewart (2008) focus on listeners’ perceptions of fillers, rather than speakers’ production, in an attempt to establish whether fillers are realised as consciously produced words. Based on a body of experimental psycholinguistic and neurolinguistic work, they claim that speakers (and listeners too) are indeed highly sensitive to the occurrence of hesitations in speech, and that on some occasions at least, fillers have the function of “heightening attention for upcoming speech” (2008: 597), consistently with their interpretation as signals. The existing evidence, however, does not justify in their view the claim that all instances of fillers are intentionally produced, as contended in Clark and Fox Tree (2002). Corley and Stewart argue that the value of fillers lies simply in “the time that they add to the utterance” and that there is little support for “the proposition that listeners are sensitive to the particular forms that hesitation take” (Corley and Stewart, 2008: 602). This would also imply that other forms of flagging, such as unfilled pauses (silence) or vowel or syllable lengthening could fulfil a very similar communicative function in speech to conventional fillers. The key factor that is common to all these phenomena is the fact that they take time to be uttered, and therefore cause a delay in speech production. It would therefore be the delay itself, rather than its specific form, which is of signalling value to the hearer. It is therefore relatively uncontroversial to claim that fillers do have some signalling value, for both speaker and hearer. Whether all instances of fillers have a signalling value, however, is quite a different matter. Kjellmer (2003) argues that “the speaker may well make use of [hesitation phenomena] unconsciously in aiming for a certain effect that s/he achieves without understanding how it was brought about” (2003: 181), similarly to what happens with intonation patterns. Kjellmer identifies several functions for fillers, including signposting of speaker turns, hesitation, highlighting and corrections, but argues that not all fillers must necessarily fulfil any of these functions.

Amongst the various functions of hesitation “highlighting” and “correction” are particularly akin to the description of flagging in CS contexts. With reference to the first, Kjellmer claims that hesitation markers can be used to obtain “a certain semantic effect” and that they function as “a sort of verbal italics, inverted commas or semantic booster” (2003: 187). The highlighting function is particularly common with infrequent words, as shown in the examples (15), (16) and (17). It should be noted that while Kjellmer ascribes a highlighting function to the examples, it would also be possible to interpret them as cases of a word search.

15. We are still at quite a **erm** rudimentary stage in terms of making use of P A S

16. That they're much more **er** discriminatory about how we use petrol

17. For six hours after the **er** aromatherapy

(Kjellmer, 2003: 187)

Highlighting can, according to Kjellmer, also function as “an appeal to the listener to accept the word or phrase in the present context despite its apparent inappropriateness” (2003: 188). In this sense, it is possible to see how the occurrence of fillers with CS mirrors and extends a strategy which is already available to monolingual speakers (of English at least). A word from a different language may be perceived as not appropriate to the context, and the speaker can use hesitation to forewarn the interlocutor – as indeed noted in Kinder (1988) and Hlavac (2006).

The second function of hesitations which is relevant to flagging is that of correction. By inserting a filler, Kjellmer argues that speakers may signal that a repair or reformulation will follow the word or phrase that has just been uttered. In a CS context, the insertion of a filler before a translation equivalent or a reformulation could also be a way of drawing the listener’s attention to the repair sequence about to begin, therefore indirectly also flagging the switched element just inserted. In studies of flagging involving CS data, neither the symptom or signal characterisation of hesitation seems to be adopted exclusively; in fact, while flags are typically seen as signals to the listener, they are also seen as symptoms of a



problem. One interpretation does not automatically exclude the other.

The interpretation of fillers is of some consequence for our understanding of flagging in CS. Even in monolingual speech, fillers or hesitation markers can signal either a delay in message delivery due to a speaker's production trouble (Clark and Fox Tree, 2002) or an element that, according to the speaker, the listener does not expect from the context (Corley and Stewart, 2008). In a bilingual context, we would argue that a switch to the other language may precisely be that type of less-predictable element for the listener. This is more likely to be the case in contexts where switches are less frequent and thus more likely to be salient, and where speakers do not habitually alternate between languages.

### 3.3.2 The notion of repair in monolingual conversation

One of the basic features of flagging which appears in almost all of the existing definitions in the literature is that in several ways it causes an interruption of the speech flow, by the insertion of an element which is not part of the speaker's main message. Another form of interruption of the speech flow that has been extensively analysed in monolingual discourse and bears on flagging is that of conversational 'repair'. The understanding of the concept of repair in conversation is to be attributed to the classic studies of interaction within the field of CA (see Schegloff *et al.*, 1977). Repair can be defined most generally as a series of "processes available to speakers through which they can deal with the problems that arise in talk" (Liddicoat, 2007: 171). Contrary to what the name implies, repair should not be seen simply as the correction or replacement of an "error", such as an inappropriate lexical item. The concept has a much wider scope, and it includes both overt and covert problems that speakers encounter during speech delivery. A word search, for instance, is an example of repair, in that it addresses a problem (also referred to as 'repairable') encountered by the speaker while formulating his or her utterance. In the case of a word search, the problem consists in the unavailability of the word itself at the moment in which it is due. An example is given in (18), in which repair occurs in the absence of an audible error. The speaker simply interrupts their utterance at *p*— and reformulates the last part of their turn.

18. L: Is his one dollar alright or should he send more than that for the p— tuh cover the postage.

(Schegloff *et al.*, 1977: 363)

In principle, therefore, nothing in speech is *a priori* excludable from the class of repairable elements (Schegloff *et al.*, 1977). Both the current speaker and their interlocutor may initiate a repair and both may also accomplish it. Repair initiated and completed by the current speaker seems to be the preferred option in spontaneous speech<sup>17</sup>.

What does get repaired may therefore be of particular significance from an interactional point of view. In CS work within CA, the presence or repair with a switch is particularly informative, as it can show the status of the switch to the listener (and the analyst too). If a switch is repaired, for instance, this may reveal that its occurrence is perceived as problematic by speakers.

Repair has also been broadly defined as “error correction, the search for a word, and the use of hesitation pauses, lexical, quasi-lexical, or non-lexical pause fillers, immediate lexical changes, false starts, and instantaneous repetitions.” (Rieger, 2003: 48). At a closer look, the list of discourse phenomena acting as instances of repair in Rieger’s definition contains many of the items which are classified as flags in the environment of a switch. Similarly to flagging, repair is therefore quite a heterogeneous phenomenon.

The study of repair strategies has predominantly focussed on to the formal and structural characteristics of the phenomenon. Repair has been defined as an orderly activity in conversation that is composed of different elements: a repairable element (overt or covert), also referred to as ‘trouble source’; the initiation of the repair; the completion/accomplishment of the repair (Schegloff *et al.*, 1977). While the repair mechanism itself can be considered a universal phenomenon, differences exist in repair strategies across different languages. Fox *et al.* (2010) show how differences in typological factors influence or determine the structure of repairs, notably repetitions and substitutions in a comparison

---

<sup>17</sup> Schegloff *et al.* (1977) identify four basic types of repair, depending on how it is initiated and achieved in interaction: self-initiated and self-completed repair; other initiated self-completed repair; other-initiated self-completed repair; other-initiated other-completed repair.

of repair strategies in English, German and Hebrew. All languages show most repetitions with function rather than content words; English shows a prevalence for recycling subject pronouns which is not shared by the other two languages, while Hebrew exhibits a higher recycling rate with prepositions. Repairs involving a substitution, on the other hand, show a uniform prevalence of content words over function words.

Language proficiency can also affect repair patterns. Rieger (2003) investigates the use of retracing – or repetitions in her own terminology – as a self-repair strategy in the spontaneous speech of German-English bilinguals (some English-dominant and some German-dominant). The speakers are recorded on separate occasions in each of their languages to ascertain whether different repair strategies are used in different languages, and the use of different strategies appears in speakers' dominant and non-dominant language. Rieger finds that speakers make greater use of fillers than retracing in their L2 compared with their L1; this is seen as evidence that fillers are an easier repair strategy than retracing (2003: 66), since their use in the L1 is lower.

Rieger's work also supports the idea that repairs are used as a way to gain time without yielding the floor, similarly to hesitation. In this way repair strategies are "delaying the production of a next item due" (Fox *et al.*, 1996: 204). This use of repairs seems to be particularly recurrent when the next item is lexical (such as a noun or a verb) rather than functional – once again consistently with hesitation. The same may also be said of several flagging strategies, notably in the case of switches to fill a lexical gap.

Interestingly, Rieger notes how on one occasion, when the time-buying strategy fails, the speaker resorts to CS to end the lexical search. This is shown in example (19).

19. June: a=ber immerhin das das **travel book** hatten wir immer als Kinder [...]

*'but after all as kids we always had the the travel book'*

(Rieger, 2003: 61)

In the example the speaker repeats the German determiner *das* 'the' as a way to repair the lexical retrieval problem she is experiencing. Presumably she could

have attempted to retrieve the German equivalent for *travel book* during the repetition of the determiner but, failing to do so, she inserts a switch to English (note that the speakers in the study had all been instructed to conduct the conversation in only one of the languages). It could therefore be argued that in this instance at least, the repair in the form of repetition is also an example of flagging. On some occasions, bilingual speakers may switch to the other language as a last resort, when they have failed to find a solution to their word search within the language they had been speaking up to that point. Rieger does not give figures relating to the occurrence of CS in her corpus, but it is possible that the example she gives is not the only token of the phenomenon in her data. In a different study of a similar population of German-English late bilingual speakers, Fehringer and Fry (2007) find that the rates of discourse phenomena produced by speakers (fillers, repetitions and reformulations) are significantly higher when informants are speaking their second language (irrespective of whether the second language is English or German) than when they are speaking in their first language. Speakers are tested separately in each language in a semi-structured interview with a mother tongue speaker of either English or German. The occurrence of increased hesitation phenomena is linked to the higher cognitive cost of speaking a second language, even though great interpersonal variation is found in the base rate of use of discourse phenomena in speakers' own L1. Regardless of this it is significant that, in spite of high proficiency, speakers produce more repairs when speaking their second language. Both studies show that there are significant differences in the production of repair (and hesitation) phenomena, for L1 vs. L2 speakers in the same language and also across languages. They also show how some differences may be exclusively attributed to syntactic and typological differences between the languages. Because of their design, however, these studies provide no information on how often speakers would produce discourse phenomena when using both languages in the same speech event, i.e. in CS.

### 3.4 Concluding remarks

In this chapter, flagging has been defined as an interruption of the speech flow occurring in the immediate proximity of an other-language insertion in conversation. A series of discourse phenomena can be classified as flags, with reference to Hlavac (2006). In investigating the existing literature it has emerged that, with few notable exceptions, flagging has received little attention in studies of CS, and its occurrence and function(s) have possibly being taken for granted. The phenomena that constitute flagging are for the most part not exclusive to bilingual speech, but occur in monolingual speech as well. A brief review of discourse phenomena occurring in monolingual speech has been undertaken, in search for meaningful similarities in the occurrence of the phenomena in the two domains. Hesitation and repair in monolingual speech appear to be closely related to flagging in CS; flagging may indeed represent a special case of functional hesitation or repair specific to the insertion of an other-language element in speech. Previous work has looked at the occurrence and distribution of hesitation and repair phenomena in monolingual speech, both by monolingual speakers and by bilinguals speaking monolingually in each of their languages. At present, however, there is no systematic account of the incidence of these phenomena in the speech of bilinguals at the points in which they are alternating between the two languages, that is, in CS. In those cases in which the monolingual counterparts of flagging have been looked at, if only to provide a baseline for a comparison with flagging of switches (see for instance Turpin, 1998), this has resulted essentially in the dismissal of flagging as a distinct phenomenon from hesitation and repair occurring in monolingual speech. The link between the two domains, and whether differences exist between monolingual and CS environments remains therefore open to further investigation. ¶

## THE DATA

**A**fter the review of previous work on flagging and on the related phenomena of hesitation and repair in chapter 3, this chapter presents the two English-Italian bilingual data sets which constitute the basis of the present study.

After introducing the general profile of speakers in both groups, the recruitment methods are explained, as well as the recording and data-gathering procedures adopted. Both practical and ethical aspects of data collection are discussed.

Of particular importance is the profile of the speakers, especially when compared with much of the existing CS literature. The present data are perhaps not of the kind typically studied, in that the speakers do not constitute a well-established community, as is often the case in naturalistic bilingualism studies. The data collected from the present groups, however, are valuable precisely because they are produced by speakers not belonging to an established community of habitual code-switchers.

In the final part of the chapter, a brief description is given of each of the recordings, in order to offer detailed information about the type of speakers, the relation existing between them and the nature of the gathered data.

## 4.1 The two data sets

The data for this study consist of two small corpora of English-Italian spontaneous conversations collected by the researcher from speakers of the two languages in Manchester and Milan. Data collection took place in the course of several recording sessions between the months of July 2008 and June 2009. Together, the two corpora amount to just under ten hours of recordings (568 minutes), comprising 271 minutes in Manchester and 297 minutes in Milan. These totals result from five Manchester conversations and seven Milan conversations. Recordings have an average length of about 45 minutes, but this varies depending on the pair of speakers, as will be detailed below (see also **Appendix 3**).

The data collection procedures adopted in this work partly replicate previous larger-scale studies carried out at Bangor University by the research team led by Deuchar as part of several on-going projects on CS in various communities. The methods used by the research group are described in detail in Deuchar *et al.* (forthcoming). These will be acknowledged throughout the chapter.

### 4.1.1 Profile of speakers

Ten speakers have been recorded in each data set. Each Manchester recording is of a different pair of speakers, and no speaker is recorded more than once. In Milan, on the other hand, three speakers are recorded more than once, although never with the same interlocutor (see **Appendix 3** for a list of the speakers in each recording). The gender split in the two groups is six males and four females in Manchester and two males and eight females in Milan. Manchester speakers have an average age of 34, and all except one were in their thirties at the time of the recording. The mean age of Milan participants is 33, although the spread is slightly wider, with four speakers in their twenties and two in their forties. Speakers in both groups are rather highly educated, with all Manchester participants and all but one of the Milan participants having university level qualifications. Several informants had or were working towards postgraduate qualifications at the time of the recording. Others were professionals or are

employed in teaching or clerical jobs. Most informants in Manchester are connected with academia to various extents.

The Manchester participants are all Italian nationals, with the exception of one who has dual Italian-British citizenship. They come from various parts of the country and speak several different regional varieties of Italian. All but one speakers claimed knowledge of additional languages besides English and Italian. Three participants reported being speakers of another Italo-Romance language as well, which was acquired early through parental input.

Participants in Milan come from various Anglophone countries – predominantly the UK and the USA. Their national identities are more diverse than that of the Manchester speakers. All informants speak their respective national standard, with some variation. Five of them also reported knowledge of languages other than English and Italian.

The informants are therefore native speakers of the language that is not the majority language<sup>18</sup> of their host country. Participants in Manchester are native speakers of Italian that have acquired English as an L2, while in Milan L1 speakers of English are recorded who have learnt Italian as an L2. Informants in both groups acquired their L2 relatively late.

Manchester speakers reported being able to converse informally in their L2 (English) while still in secondary school. The widespread presence of English as a subject in the Italian school curriculum seems to be one of the reasons for this. The majority of Milan informants, on the other hand, reported acquiring significant skills in their L2 (Italian) only in adulthood, either through higher formal education or after moving to Italy. In terms of self-rated proficiency, six out of ten speakers in each group report to be “confident in extended conversation” when speaking their L2 (the highest score on a four-point scale). The others reported being on the step immediately below that, i.e. “fairly confident in extended conversation”. Only one of the Milan participants reported being confident only in basic conversation.

---

<sup>18</sup> With the exception of speaker Daisy (MilaH), who is a native speaker of Italian (see description of recording **Milano 8** in 4.4.1). Daisy does not produce any switches for the duration of the recording, meaning that *de facto* she supplies no data for the present study. She has not been completely excluded from the study since she is Kelly’s interlocutor in **Milano 8**.



Most speakers in both groups report being educated through their L1 in both primary and secondary school. Manchester informants report receiving further education in either Italian (six out of ten) or both English and Italian (four out of ten). Milan speakers, on the other hand, overwhelmingly report receiving English-medium higher education (eight out of ten).

While there are differences between the groups – most notably in terms of age of acquisition of the L2, the overall picture that emerges is that of speakers with comparable features. Individual profiles of speakers are presented in table form in **Appendix 5**. Information about the speakers as groups is presented graphically in **Appendix 6** (Manchester) and **Appendix 7** (Milan).

## 4.2 Selection of groups

The original aim of the project was to compare two bilingual communities sharing the same pair of languages (English and Italian) but with different profiles as far as their language proficiency is concerned. It was therefore decided to study two groups of bilingual English-Italian speakers in similar environments. Comparable settings were chosen in order to reduce as much as possible the number of external factors that could affect a variation of speakers' language-mixing patterns. The rationale for attempting to control for this was to allow an assessment of the respective role of linguistic and extralinguistic variables in language mixing patterns. Eventually, suitably comparable speakers were identified in the cities of Manchester and Milan. These cities were chosen for several reasons: the two cities are comparable in terms of population and general socio-economic context. Both cities are at the centre of large post-industrial conurbations and are home to large immigrant communities and sizeable international student populations. On a more practical level, the two locations were also relatively convenient as recording locations for the researcher.

With regard to the groups under study in the present work, Manchester has the second-largest Italian presence in the United Kingdom after London (an estimated 25,000 people in the Greater Manchester area)<sup>19</sup>. The Milan area is home to a

---

<sup>19</sup> Sources: <http://news.bbc.co.uk/1/hi/england/manchester/3223776.stm>; <http://www.easymilano.it>. Last accessed in August 2011.

large international community, and while official figures are not easy to obtain it is estimated that at least 30,000 UK, USA, Canada, Australia and New Zealand nationals live in the city of Milan and the surrounding area<sup>20</sup>. To these, smaller numbers of citizens from other Commonwealth countries should be added, as well as citizens of other countries who don't speak English as their mother tongue, but who use it routinely in their day to day life as members of the international expatriate community.

#### 4.2.1 Definition of bilinguals

The term 'bilingual' is used here in a broad sense. Members of the two groups under study could be described more precisely either as 'late bilinguals' or 'highly-proficient L2 speakers', although a certain degree of variation can be noted in individual levels of proficiency – both self-reports acquired from participant questionnaires (see **Appendix 4**) and the researcher's own observations during interaction with participants and debriefing. The range of variability in proficiency seems to be somewhat broader in Milan, where L2 speakers of Italian range from the intermediate learner to the native-like. Manchester participants, on the other hand, are more uniform in the command of their L2, which is high but not native-like, at least at the phonological and prosodic level.

As already noted, in both groups acquisition of the L2 started at school age or even later, in adulthood. Two participants – one in Manchester and one in Milan – described themselves as being simultaneous English-Italian bilinguals. Both were born from parents with different L1's and reported being exposed to both English and Italian from a very early age. In no case did participants report growing up or being exposed to English-Italian bilingualism at the level of the community during their upbringing.

---

<sup>20</sup> Source: British Consulate-General in Milan (personal communication, January 2011).

#### 4.2.2 Types of groups in code-switching research

Previous work on CS has looked at patterns of language mixing in very different groups of speakers. A first category is that of relatively stable bilingual or multilingual areas. Well-known examples in the literature of this kind of populations include Spanish-English speakers in New York (Poplack, 1980) or Welsh-English speakers in Wales (Deuchar and Davies, 2009). A second type is that of multilingual communities in postcolonial settings (see for instance Myers-Scotton, 1993b). In some of these groups a rapid change in the balance between the languages may be underway (see the case of French-Moroccan Arabic in Bentahila and Davies, 1995). A different strand of communities that have sparked the interest of scholars has to do with recent migratory movements, in situations where language shift may be underway and new linguistic minorities come into being. Examples of this kind of study are found in the work of Backus (1996) on the Turkish community in the Netherlands, Clyne (2003) on a number of immigrant languages in contact with English in Australia and Li Wei (1994) on Chinese-English in the UK. In Europe at least, this phenomenon has sometimes been characterised as “urban multilingualism” (Extra and Yağmur, 2004). More often than not, the populations investigated in these studies are described as being part of rather close-knit communities. The members of these groups typically show high-density social networks, and are frequently also concentrated geographically. Speakers are also aware of their status as members of a certain community or exhibit strategies of marking their allegiance to the group. This allegiance is not always but frequently constructed in relation to the ideas of ethnicity, nationality and language or a combination of these elements. The present study is different from those referred to above in the choice of the populations under investigation and their features. Participants in both groups are first generation immigrants in the respective host countries. All participants are under fifty years of age and most of them have moved to their host country as young adults. Their length of stay in the host country is generally under ten years. Participants in both groups are rather scattered geographically – while some live in the inner city, others are spread in the surrounding hinterland. There is no concentration in particular areas or neighbourhoods.

In terms of their linguistic and socioeconomic profiles, participants could arguably be compared to the L2 and L3 learner populations investigated in Poulisse and Bongaerts (1994) and Dewaele (2001) or the “unbalanced bilinguals” analysed in Lüdi (2003), some of whom were also learners. In these groups the age and type of acquisition are more similar to the speakers under analysis in the present work than in most of the aforementioned studies.

There are, however, some fundamental differences between the populations of language learners and the present study. In all the works on learners mentioned above, speakers (whether in experimental tasks, classroom settings or informal interviews) consistently switch from their L2 (or L3) into their L1. Speakers in Manchester and Milan, on the other hand, consistently switch from their L1 into their L2 (see chapter 6).

CS in language learners is often understood as a coping process due to lack of proficiency in the L2 – a mechanism defined by Lüdi as ‘translinguistic wording’ (2003: 177). This involves the speakers resorting to their strongest language while using the L2, in order to avoid a breakdown in communication. The incidence of this phenomenon is expected to be higher when speakers are interacting with a fellow learner or a bilingual interlocutor, rather than a monolingual speaker of their L2. Switching from the L2 to the L1 has also been interpreted as unintentional interference, due to the higher psycholinguistic activation of L1 forms (especially in the case of function words, as shown in Poulisse and Bongaerts, 1994: 53).

In the present data, besides the difference in directionality of switching compared with other studies, switches mostly involve content words (see the analysis in chapter 6). Arguably, the difference in directionality may be explained in terms of methodology; the CS studies on learners involved semi-structured interaction with a (sometimes bilingual) native speaker of the learners’ L2, effectively constraining the language of interaction. In the present data, by contrast, dyads of speakers with similar linguistic profiles were recorded having a free conversation.

The prevalence of the speakers’ L1 in what was effectively an in-group situation should therefore not be seen as surprising.

There are other important features of the Manchester and Milan groups that

should be highlighted. Participants' social networks are made up mostly of individuals who are members of the linguistic majority, i.e. Italians in Milan and English-speakers in Manchester. In questionnaires, speakers were asked to list their five closest contacts and the languages they normally use with them (English, Italian, English and Italian or other). From this information a score was calculated to capture the overall orientation of speakers' network in terms of language use (see **Appendices 6** and **7**). Both groups seem to be somewhat oriented more towards the linguistic majority (i.e. their L2) than their L1, as far as their language choices with their most frequent contacts are concerned. This orientation seems to be more marked in Milan than in Manchester.

Further evidence for the more marked orientation to the majority language by Milan informants is to be seen in the observation – again both from questionnaire data and direct participant observation – that many speakers are in relationships with or married to non-members of their respective groups. The projection of the speakers' social network outside the groups also seems to be facilitated by their work environment, through which they associate both with fellow members of the group and non-members. It could therefore be argued that speakers are quite well integrated into the mainstream society of their host country.

### 4.2.3 **Groups or communities?**

When all of the features of speakers presented in **4.2.2** are compared with those of the populations in the studies mentioned earlier (perhaps with the exclusions of studies on L2 learners), the question arises of whether the groups under study here can in fact be said to constitute a community. While conceding that the definition is much disputed in social sciences, Milroy (1987: 14) notes that amongst the factors that define a community are:

- a. The consciousness of belonging to a cohesive group;
- b. The association to a strong territorial basis or “localism”.

Milroy characterises the notion of localism as typically involving two factors: the spatial concentration of group members, and the type of social interactions that they engage in. In close-knit communities, she argues, interactions tend to be for the most part within the neighbourhood in which group members are concentrated. The neighbourhood itself comes to be seen as an extension of the individual's home, and an integral part of his/her immediate social network. This narrow definition of 'community' seems to be unsuitable in the present situation. It does not adequately capture the social dynamics of the two populations, given the openness of their social networks, (see **Appendix 2** for a graphic representation of the relationship between speakers), their relative lack of attachment to particular areas or neighbourhoods and the apparent lack of any institutionalised form of aggregation or association.

A broader way of characterising these two groups may be the idea of 'speech community' (see Saville-Troike, 2003 for a review). Saville-Troike observes that while "the essential criterion for "community" is that some significant dimension of experience be shared" for a speech community it is essential "that the shared dimension be related to ways in which members of the group use, value, or interpret language" (2003: 15).

Speakers in the two groups have a largely shared linguistic repertoire, all share the experience of being expatriates, and to an extent can be shown to be constructing norms for the interpretation of their speech. They could perhaps be seen as incipient or marginal speech communities, given that the norms of interaction among members (particularly with reference to the use of two languages) are not stable and are not central to the speakers' socialisation patterns. This is shown, for instance, by the orientation of the speakers' social network to members of the linguistic majority, already mentioned in section **4.2.2**. For the purposes of the present work, speakers will simply be characterised as 'groups' or 'clusters'. The word 'community' may be occasionally used, but this has to be interpreted in the loosest possible way, and does not imply conformity of the groups to the defining characteristic of a community cited above.

#### 4.2.4 Recruitment of speakers

In the early stages of the study it was decided to recruit balanced or simultaneous bilinguals as much as possible. The reason for doing so is that there is a large body of literature dealing with this type of populations, which would have made comparisons with previous research and existing data easier. It soon became apparent, however, that finding sufficient numbers of speakers fulfilling this profile would not be viable with the resources available. In the absence of an established or stable bilingual community, balanced bilinguals turned out to be quite rare. Furthermore, they often would not be acquainted with many other balanced bilinguals, meaning that they would not have a friend or acquaintance with whom they could participate in the study. For these reasons, it was decided to adopt less strict criteria for recruiting speakers, and adopt the more operational definition of bilingual outlined earlier.

Several methods of recruiting participants were explored for the study. Initially, advertising on websites and media targeted at the two groups (such as specialised news bulletins or classified adverts magazines) was considered. The idea was subsequently discarded due both to budgetary reasons and reservations about the response rates to be expected, since it was not possible to offer a substantial incentive for participation. Several groups and organisations for expatriates were approached, such as reading clubs, religious congregations, language schools and other associations. The approach took the form of a request to a contact person in the organisation for collaboration, to whom the general aims of the project and the type of participants required were explained. It was hoped that these organisations could act as first points of contact with potential participants. These attempts, however, generated little or no response, and were also abandoned. Eventually, the most fruitful procedure for recruiting participants proved to be the oft-cited “friend of a friend” approach (Milroy, 1987: 53; 1989: 66). This method involves the researcher being introduced to potential participants through a common friend or acquaintance. The method has many advantages; the main one noted while collecting data for this project is probably that potential informants are much more willing to participate or help recruiting when the researcher is directly introduced to participants by a friend.

The exact dynamics of recruitment, however, were slightly different for the two groups. In Manchester, an Italian acquaintance of the researcher who happened to be living in Manchester contacted a list of her own friends as potential participants. It should be noted that many of these friends had some involvement in academia (i.e. they were postgraduate students or junior researchers), while others were Italian professionals based in or around Manchester. Some of them had already taken part in surveys and/or academic studies. Participants were contacted by email and were invited to respond directly to the researcher. Respondents were then informed about the study more in detail by the researcher and appointments were made for the recording session. In Milan, recruitment started from a single contact. A friend of the researcher was at the time working in an educational institution which employs several British and American nationals. One of these was introduced by the contact person to the researcher and agreed to participate as well as help with recruitment of participants. After initial contact was made with the first participant, it was easier to recruit more informants. The first participant's social network proved to be invaluable in this sense. She and one of her contacts provided several contacts, as well as appearing in more than one recording, each time with different partners. In both groups appointments were made both by phone and email. It should be noted that participants were not offered money for taking part in the study.

#### 4.2.5 Briefing of participants

Before meeting for the recording, participants were provided with some basic information about the aims of the study. When participants were contacted by email only, the information was given to them in the form of a bilingual document (see **Appendix 1**), introducing the research and the general information about the project. While participants were not informed that CS or language alternation would be the main focus of the research, they were told that the project was about communication between bilinguals, simply defined as “people who use English and Italian on an everyday basis”. It was also mentioned what they would be expected to do as part of the project and that explicit



authorisation would be required from them in order for the researcher to be able to make use of the data acquired. They were also informed that their anonymity would be protected.

All participants were asked to be recorded while having an informal conversation with a bilingual person of their choice, who could be a friend, relative or colleague. When they agreed to take part in the study, they were also asked to think of somebody they would like to be recorded with. They were told that the interlocutor should be somebody that participants are comfortable talking with. When making arrangements with participants, the researcher tried to leave the choice of the language of interaction up to speakers as much as possible. In Manchester, it was observed that Italian was the preferred choice for interaction with fellow Italians. Being Italian himself and living in Britain, the researcher was perceived by participants as being part of the same group of speakers and treated accordingly. As a matter of fact, none of the Manchester participants selected English as the language of interaction when communicating with the researcher by email, phone, or during face-to-face interaction.

The situation in Milan was more variable. The original contact person expressed a preference for Italian in the earlier stages, but she then spoke either language in subsequent meetings. The researcher observed that the presence in the venue of extra people or other participants had a stronger impact on language choice than it had in Manchester. It should be noted that Milan participants themselves seemed at times hesitant about what language to use with the researcher. Sometimes they asked him which language he would rather use with them, to which he always responded in a way that left their options open. This uncertainty was perhaps due to the researcher's ambiguous position as an Italian coming from a British institution. On several occasions in Milan it was noted that participants spoke Italian to the researcher before the recording but then used mostly English – the language they had been using predominantly during the recording – after the recording and in the de-briefing phase.

Participants were also asked to choose where they wanted the recording to take place: workplace, home or cafés were the most frequent choices.

The rationale for giving participants the task of choosing the venue as well

as their partner for the conversation was to increase the chances of achieving informal and relaxed speech during the recording; on the one hand the researcher reduces his control of the setting, but on the other he can ensure that speakers will not feel intimidated by their surroundings. The only requirement given was that the locations were not excessively noisy or crowded, in order to guarantee adequate recording quality. This method of recruiting the second participant also constitutes a *sui generis* instantiation of the friend of a friend approach, since it is expected that speakers will normally choose members of their immediate social network.

### 4.3 Ethical issues and speaker consent

When collecting naturalistic language data, researchers need to find a compromise between two opposing requirements: on the one hand the wish for data that reflects spontaneous language use as closely as possible; on the other the need to gather data ethically and be transparent with participants.

These methodological issues are linked to what is commonly referred to in the literature as the Observer's Paradox (Labov, 1972), which can be formulated as the need to systematically observe people's behaviour while they are not aware of being observed. When looking at CS, this becomes particularly important: in many groups CS is a somewhat stigmatised behaviour; it is therefore paramount that speakers are not aware of what the investigator is looking at, since this could increase awareness of their own speech production and reduce the occurrence of phenomena – such as CS – which are often perceived as belonging to more casual speech registers (Poplack, 1982).

Ideally, therefore, researchers should collect 'candid' data from participants who are unaware of being recorded at least until after the recording has taken place. In this way, the presence of a recorder would have no effect on informants' speech patterns. In order to be able to gain access to informants' speech surreptitiously, researchers should be able to find speakers that are concentrated in a precise area or neighbourhood, and who frequently interact with members of their own group. Geographical concentration is a feature of several bilingual communities, such as

the Puerto Rican Spanish-English community analysed in Poplack (1980). Such a community could, theoretically, make candid recording a practical option. Candid recording, however, is not an ethically satisfactory conduct. It is normally required that participants are fully aware of being part of a research project before data have been collected from them. Full disclosure after data collection has taken place, which has in the past been used to obtain candid data in the social sciences, is also an unacceptable behaviour. Participants may feel that their rights and dignity have not been respected when data have been acquired for them without their explicit consent. Ethical consideration alone therefore categorically exclude the possibility of surreptitious recording, even before the more practical difficulties that it would entail are taken into account. Unlike some of these large-scale studies in which a particular geographically-concentrated community is investigated and participants could be easily recruited, the present work only aimed to recruit speakers with very precise characteristics. As seen in section **4.2.3**, potential participants in Manchester and Milan are not members of a close-knit community, nor could they be found in an easily circumscribable area. As observed earlier, they tend instead to be scattered over a relatively large area and to spend a significant part of their time interacting with monolinguals or non-members of the target group. It would therefore be very difficult for the researcher to obtain candid recordings, since members of the groups are not easy to find and the chances of finding them while engaged in conversation with another in-group member are fairly low. Since recording sessions with participants were arranged in advance, they were given detailed information on how the collected data would be handled and used by the researcher or other subjects. From the first contact, they were told that they would be recorded talking in pairs while the researcher was not present. Participants were also informed that the data would be transcribed by the same person doing the recording and that the both recording and transcript could be eventually be made available on the Internet. The transcriptions of the data would be completely anonymous, with names of speakers and any individuals mentioned replaced with pseudonyms – with the exception of public figures or fictional characters. Names of organisations, companies or other institutions

mentioned in speech would also be replaced with synonyms when they could potentially lead to identification of the participant or when sensitive information was being disclosed. These measures were explained verbally to participants when arranging an appointment with them and also just before the recording took place. It was believed that ensuring that participants were aware of the anonymity of the recording would contribute to make them feel less self-conscious about their speech and encourage them to talk more naturally. After the recording session, all participants were asked to sign a consent form which detailed how the data would be used exclusively for academic and/or teaching purposes and how it could be shared and made available within the academic community and on the Internet (while always protecting speakers' anonymity). The consent form was made available to speakers in both English and Italian, and also asked participants for their contact details, should the necessity arise to get back to them at a later stage. The researcher also made sure that his own contact details would be available to participants, in case they decided to opt out of the study or withdraw their consent to the data being used or made publicly available at a later time.

Participants were also given the opportunity to listen to all or part of their conversation after the recording sessions and were asked whether they would like any portions of the data to be edited out of the final version. None of the participants, however, asked for sections of the conversation to be removed from the recording or to listen to the recording.

In one instance only (**Manchester 1**) one of the speakers refused to give full consent to the use of the data. It was then agreed that the transcript of the recording would be available with no extra restrictions, while the audio file would not be made available online.

An additional ethical issue to be dealt with was that of third parties appearing during the recording, i.e. speakers other than the two participants joining in the conversation. This happened both in public spaces and in participants' homes. Often these extra speakers' incursions were quite brief and terminated as soon as they realised that participants were being recorded. In other instances, however, third parties, while not greatly involved in the main conversation, were present

for extended periods and intervened several times. This is particularly the case in one recording (**Milano 8**). Although their utterances were often audible, no consent has been obtained for these speakers. These speakers' utterances have therefore not been transcribed. When one of the participants is addressing an external speaker, this is normally noted in a comment in the main body of the transcription. In this way, some information is inevitably lost.

An alternative solution to this problem would have been to request consent for these additional speakers. These speakers, however, sometimes happened to be monolinguals, i.e. not part of the group under study. Also, the researcher often realised that extra participants were present in the recording only when listening to the conversation at a later time. Excluding extra participants from the analysis was therefore adopted as the most practical and ethical solution.

#### **4.4 Recording procedure**

As mentioned above, appointments were normally arranged by either phone or email. At this point, the researcher would normally only have direct contact with one of the participants, since the other would be chosen and contacted by the first participants themselves. On some occasions this caused difficulties. While the profile of the ideal interlocutor was always explained explicitly to the participant who was approached directly, participants sometimes did not understand the exact specifications required and invited individuals who were not ideal for the study. This potential risk notwithstanding, leaving participants to select their conversation partner was opted for as the simplest and most effective way to ensure a more naturalistic recording.

The researcher would meet both participants at the pre-arranged location. He would briefly explain the general aims of the project, mostly for the benefit of the second participant. In most cases, the researcher would not meet the second participant until the recording session itself, since the first participant was responsible for recruiting them.

While illustrating the project to the participants, the researcher would also set up the recording equipment and do a trial recording. Participants were invited to

make themselves comfortable – usually the two speakers would be sitting across a table, whether in a public space or private house. It was explained to them that they would be free to move around during the session if they so wished, provided that they took the recording equipment with them.

All recordings were made using a M-Audio Microtrack digital recorder. Two long lead lapel microphones were connected to the right and left channel of the recorder. The reason for using two separate microphones is that analysis and transcription are made easier, since it is possible to listen to only one of the two audio channels. This is particularly useful in the case of overlaps. Using a single built-in microphone would have made the equipment less obtrusive (lapel microphones had wires and were clipped onto participants' clothing), but would have not allowed the analyst to listen to the two speakers separately. It should be noted, however, that in some recordings one or both microphones picked up both voices, especially when speakers were talking loudly or if they were sitting quite close to one another.

Participants were not given explicit instructions on how to conduct the conversation other than being invited to have an informal conversation for a certain period of time. The researcher would suggest between 30 and 45 minutes. In some cases speakers would only be available for a shorter period of time, while in others they were happy to talk for longer. The length of the recordings therefore varies. No list of topics or other explicit instructions were given, although when participants were colleagues or working partners, they were told that they could make use of the recording time to discuss issues pertaining to their job if they wished to do so.

Participants were also told that the researcher would not be present during the recording. This seemed to surprise several participants, who had expected the recording session to be a more structured interview conducted by the researcher. Sometimes speakers would ask in what language they should conduct the conversation, either English or Italian. The response that was normally given was that it did not matter, as long as they talked in the way they normally did with one another. No explicit mention of language mixing or CS was ever made at this stage by the researcher.

After giving the instructions, the researcher would normally leave the venue, only

to come back after about 45 minutes. In some cases, speakers specified in advance that they would only be available for a shorter period of time. Their preferences were generally accommodated so as not to lose potentially valuable data.

During the recording the researcher was not able to monitor the speakers from a distance. In one session (**Milano 1**), unbeknown to the participants, the equipment stopped working ten minutes into the recording due to a technical problem.

The file was subsequently excluded from the analysis for being too short.

Upon his return, the researcher would first stop the recorder and check that the audio file had been saved correctly and would give speakers consent forms and questionnaire (see **Appendix 4**) to fill in. The questionnaire contains general information on speakers' socioeconomic status, language background, proficiency self-rating, patterns of language usage and language attitudes and is modelled after the one used for the Siarad corpus and related projects. Speakers were also asked whether they had encountered any difficulties in keeping the conversation going or whether the presence of the recording equipment had made them feel uncomfortable. Most speakers reported that while they expected to be somewhat intimidated by the situation, or to fail to speak for the duration of the recording, they were quite at ease once they started talking.

When they asked explicitly and when the researcher knew that participants would not take part in further recordings, the exact aim of the study was revealed. In most cases, speakers did not seem to have recognised previously the aims of the project – they normally thought that the researcher would be interested in the actual topics of the conversation. In two of the Manchester sessions, participants were postgraduate students working in language-related areas. One of them in particular (SAR in **Manchester 1**) was familiar with the concept of code-switching.

#### 4.4.1 **Description of recordings**

In this section, a brief description is given of each recording. Participants' age and gender are given in brackets next to their names (which are fictitious) below each recording heading.

The relationship between speakers is sketched, as well as the setting in which

the recording took place and a summary of the main topics covered. Other observations are also included, about speakers' perceived attitudes, spontaneity and any awareness they may have had about the aims of the study. As noted earlier, three participants in Milan appear more than once. Kelly appears three times (**Milano 2, 4, 8**) with different partners; Emma appears in **Milano 2** and **Milano 6** and Jonathan is recorded in **Milano 5** and **6**. In Manchester, no participant features in more than one recording (see **Appendices 2** and **3**).

## **Manchester 1**

Participants: Antonio (41 M), Sara (35 F)

Sara is finishing her doctorate at the University of Manchester, while Antonio is an architect. He works in an English-speaking environment and does not seem to have many Italian acquaintances, unlike Sara. Both have lived in the UK for over five years at the time of the recording. They both come from the same city in the North East of Italy but only met in the UK. They used to share a house in Manchester and are well acquainted with one another. The recording takes place in Sara's living room early on a Saturday afternoon. The two talk mostly about relationships, their respective jobs, common friends and plans for the future. Switches are relatively frequent, especially from Antonio. Antonio is not always a very co-operative interlocutor, which may have been due to his not feeling very well at the time of the recording. He claims to be tired and decides unilaterally to interrupt the recording a few minutes before the researcher's return by removing his microphone and going to a different room.

## **Manchester 2**

Participants: Paolo (33 M), Luca (35 M)

Paolo is from the South of Italy but has been living intermittently in the UK in his adult life as part of his job as a translator. He has relocated to Manchester on a semi-permanent basis less than a year before the recording took place. He works as a translator and an Italian tutor and is attempting a career change. Luca comes



from the North of Italy and has relocated to the UK for study reasons. He has been there for about three years. He is a colleague of Paolo's and a postgraduate student. The recording takes place in a campus café on a Friday afternoon. In the first part of the recording Paolo asks Luca about opportunities for employment or further research in the UK and they compare the academic systems in the UK and Italy. A third party (Italian speaker) joins the discussion at some point but he leaves quite quickly. Luca and Paolo subsequently move on to talk about more personal issues and speak extensively about a common acquaintance. The researcher finds them still talking. They report not being intimidated by the recorder but having spoken extensively and freely.

### **Manchester 3**

Participants: Giulia (30 F), Marina (30 F)

Giulia is the only English-Italian balanced bilingual in the Manchester data. She has an English mother and an Italian father and was born and brought up in central Italy. She is a postgraduate student and has lived in Manchester for five years at the time of the recording. She relocated to the UK upon leaving school in Italy. Marina is a postgraduate student at a different Manchester institution. She comes from the north of Italy and has spent the past two years in Manchester. She moved to Manchester due to the strong links between her current and previous university. The two speakers are close friends and share many acquaintances. The recording takes place on a Friday evening in a café in central Manchester. The two speakers mostly discuss private matters, such as relationships and common friendships. They also start making arrangements to celebrate their birthdays together. Towards the end of the recording Marina receives a phone call which causes her to switch completely to English for the duration of the call. Both speakers seem to be quite oblivious of the recording equipment. When the researcher returns, they both appear at ease. Neither seems to have any ideas about the aims of the study.

## **Manchester 4**

Participants: Nicolò (32 M), Giorgio (33 M)

Nicolò and Giorgio are both postgraduate students in the same department in Manchester. Giorgio was born and brought up in the North of Italy and has been living in various locations in the UK and Ireland for almost ten years.

He has lived in Manchester for three years at the time of the recording. Besides his studies, he also works as an Italian language teacher. Giorgio was born in Sardinia and has lived in Manchester for about two years. The recording takes place in a café in central Manchester on a Saturday afternoon.

After talking about mutual friends briefly, the two speakers extensively discuss issues pertaining to their research. At first they both seem quite uncomfortable with the setting, but they appear to relax and talk more casually as they become more engaged with the subject matter. They have some ideas about the aims of the study, which may have had an influence on their speech.

## **Manchester 5**

Participants: Vittoria (34 F), Gavino (35 M)

Vittoria and Gavino are both postgraduate students in the same department in Manchester. Vittoria was born and raised in the North of Italy but has also lived in various locations in Central and South America as an adult. She has been in Manchester intermittently for over two years. Gavino was born and raised in Sardinia and he also spent time in South America before moving to the UK, where he has been residing for almost six years at the time of the recording.

The conversation takes place in Gavino's house. Since the two speakers work in the same department, they decide to spend the time talking about their research. Gavino is working on an academic project and asks for Vittoria's advice. The two are sitting in front of a computer looking at various short video recordings and they discuss how to present the data. They both seem quite engrossed in their task and completely unaware of being recorded. They do not report any difficulty to the researcher at the end of the session.

## **Milano 1**

This recording was unexpectedly short due to a problem with the recording equipment. The data gathered and information about the speakers have therefore been excluded from the analysis.

## **Milano 2**

Participants: Emma (36 F), Kelly (24 F)

Emma and Kelly used to be colleagues in a higher-education institution in Milan. Emma is British and has lived in Milan for two years. She is in a long-term relationship with an Italian. Her job involves pastoral and administrative duties in relation to students. Kelly is American and after leaving the job she shared with Emma she has been working as a tutor, giving private English lessons to both children and adults. She has lived in Italy for three years and is married to an Italian. The recording takes place in Emma's workplace on a Friday afternoon. The two alternate between discussing work-related problems and talking about common acquaintances. At times, some Italian colleagues of Emma's enter the room and interact with speakers, but their incursions are very brief. These cause both speakers to shift to Italian when addressing them. Kelly has also brought her own dog with her. She addresses it in both languages.

Speakers are well acquainted and sound comfortable during the recording. Emma is working at a computer while talking and appears distracted at times.

## **Milano 3**

Participants: Kylie (41 F), Amiria (35 F)

Kylie is Scottish and works in the media industry. She has lived in Milan for three years, is married to an Italian man and has two children. Amiria is a New Zealander who was born in the United States. She studied Italian at University and has lived in Milan for over ten years. She works as a primary school teacher, is also married to an Italian and has two children. The two women are friends and met through a common acquaintance. The recording takes place in Amiria's

flat on a weekday evening. The two women are sitting at a table while Amiria is feeding her one-year-old child. Her husband is in the flat and occasionally joins in the conversation, using both English and Italian.

The conversation is mostly about school and education: the two women talk extensively about the Italian school system and express their concerns about English language provision for their children. They also discuss plans for family activities and holidays. Both women also address Amiria's baby from time to time, always in English. When the researcher returns the speakers seem to be at ease and engrossed in the conversation.

## **Milano 4**

Participants: Kelly (24 F), Anna (23 F)

Kelly already appears in **Milano 2**. Anna is the only balanced bilingual in the Milan data. Her father is Italian while her mother is American. She was born and lived in the United States until the age of eight, after which she moved to Italy. She commutes to Milan for work and study. The speakers are close friends and co-workers: they both give private English lessons, some of which are arranged through an agency for which they both work. The recording takes place in Kelly's flat on a weekday afternoon. Kelly is organising teaching materials during the recording as she is expecting a student in the evening. The tone of the conversation is extremely informal. The conversation mostly revolves around the speakers' work as private tutors. A lot of time is spent commenting on the families that they work with and their lifestyle. Both speakers appear very at ease talking to one another and completely unaware of the recorder.

## **Milano 5**

Participants: Jonathan (26 M), Emily (38 F)

Jonathan is an American postgraduate student at a university in Milan. He has lived in Milan for one and a half years and he is in a relationship with an Italian girl. He is of Italo-American descent but did not get exposure to the language

through his family. Emily is English and is a secretary and support officer in the same institution. She has lived in Milan for about seven years, but had previously spent a year in Italy after graduating. She is married to an Italian. The relationship between the speakers is strictly professional, but they are quite well acquainted with one another. The recording takes place in the university's staff room on a weekday morning. The session has to be kept short due to Emily's commitments. Topics discussed include the difficulty of learning a new language and relationship problems with partners from a different background. Emily briefly addresses some staff members in the room and switches to Italian when doing so.

## **Milano 6**

Participants: Jonathan (26 M), Emma (36 F)

Both Jonathan and Emma have featured in previous recordings (**Milano 5** and **2** respectively). The session takes place in the same venue as **Milano 5** (a university staff room), just after the previous recording has finished. Jonathan mentions to Emma some of the topics discussed with Emily, and the two continue along the same line. They also discuss their respective personal lives and the language learning process. They also spend some time talking about the increasing presence of the English language in all spheres of Italian life and how this can sometime generate confusion. The two speakers seem quite at ease with one another. Emma is very conscious of being recorded towards the beginning of the session and initially provides minimal responses, but she becomes more relaxed after a few minutes.

## **Milano 7**

Participants: Alistair (47 M), Fiona (40 F)

Alistair is a British IT consultant working for a multinational corporation in Milan. He is married to an Italian whom he met in the UK and has two young children. He has lived in Milan for nine years. Fiona is an American of Italian descent. She is the owner and director of a nursery. She is married to an Italian

and has two children. She has lived continuously in Milan for fifteen years. The recording takes place in Alistair's flat one weekday evening. The two speakers are neighbours. They have known each other for years but are not close friends. The beginning of the recording is characterised by small talk and seems quite awkward, but speakers become comfortable with the setting quickly. They mostly talk about their personal histories and how they came to meet their spouses, as well as the difficulty of leaving families behind in their home countries. Alistair's wife makes a brief intervention towards the end of the session but is otherwise absent from the recording.

## **Milano 8**

Participants: Daisy (26 F), Kelly (24 F)

This recording differs from the previous ones in that one of the speakers, Daisy, is a L1 speaker of Italian who has lived in the US for a year during her adolescence as part of a school immersion programme. She speaks English as an L2, although she is highly proficient. Her interlocutor, Kelly already features in **Milano 2** and **4**. The recording takes place in Kelly's flat on a weekday evening. Kelly has invited Daisy to her house for dinner. Their respective partners – both L1 speakers of Italian – are also present. The two speakers appear to be close friends, and do not seem intimidated by the presence of recording equipment. The conversation is very informal, with topics ranging from work-related issues to weekend plans and common acquaintances. The setting also differs from other recordings since the speakers' partners are present for a considerable part of the session. Their interventions are quite limited and tend to be in English. The impression is that English is also the preferred language of interaction between Daisy and Kelly.

## 4.5 Concluding remarks

This chapter has introduced the data for the analysis, and provided details of the data-collection procedures adopted. The methods for recruitment of participants and the general features of speakers in the two groups have also been presented. From the description of the Manchester and Milan groups it emerges that the speakers under analysis do not constitute a community, in which identifiable speech patterns and norms tend to be expected. The two groups could perhaps be seen as an incipient speech community, in which norms have not yet crystallised.

The value of CS data from such groups, however, lies precisely in the absence of a strongly identifiable community dimension, in that the emergence of bilingual discourse practices can be observed even in the absence of CS as an established “discourse mode” (Poplack, 1988; Eppler 1994) in the speakers’ repertoire.

The final part of the chapter describes the procedures for the recording of conversations. The single recordings that make up the Manchester and Milan data sets have also been described. Additional information about individual speakers has been provided for each recording based on direct observation and participants’ questionnaires (see also **Appendices 6** and **7** for questionnaire results) together with a brief description of the setting, topics covered and relationship between speakers in each recording. ¶

# METHODOLOGY

The definition of flagging adopted in the present work is introduced in this chapter. Examples of different types of phenomena falling under the definition of flagging are presented with examples taken from the Manchester and Milan data.

Subsequently, the transcription of the data and the theoretical principles underlying the coding of spontaneous speech are discussed. The CHAT system, which is used to transcribe the present data is introduced, together with an illustration of additional transcription conventions adopted. Particular attention is given to the problem of assigning a language tag to single words, and the criteria adopted when dealing with ambiguous cases.

## 5.1 Definition of flagging in the present work

As noted in section 3.4, flagging is defined as an interruption of the speech flow occurring in the proximity of an other-language insertion. This interruption can take several forms, and be achieved by a range of phenomena typical of spoken speech. The list of discourse phenomena classed as flagging in the present work is based largely on Hlavac's (2006) typology of flags, which has been adapted for the English-Italian data. All discourse phenomena occurring in the immediate proximity of an other-language element, either before or after the element itself, are classified as flags. It follows that an other-language segment which appears in the data together with any of these phenomena will be classed as a "flagged switch".



The definition of flagging adopted is purely operational and has been used when extracting data for the analysis. It should be noted that, unlike in Poplack's work, there is no necessary *a priori* assumption about the discourse or pragmatic function fulfilled by the other-language element, or about the value of the flag itself. The list of phenomena classed as flags when occurring in the vicinity of a switch in the present data is reproduced below.

**Phenomena preceding or following the switch (flags)**

- i. Paralinguistic markers (e.g. laughter, nervous coughing)
- ii. Unfilled pauses (silence)
- iii. Lexicalised filled pauses (*like, you know*)
- iv. Non-lexicalised filled pauses (*uh, uhm*)
- v. Other-language equivalents
- vi. Explicit pre-empting (pre switch only)
- vii. Explicit justification (post switch only)
- viii. Disfluencies, false starts, retracing
- ix. Vowel/syllable lengthening.

The category of vowel/syllable lengthening in (ix) has been added to Hlavac's original typology. It was noted in the data collection and transcription phase that speakers in Manchester and, to a lesser extent also in Milan, often seem to slow down their speech rate or lengthen the syllables in the immediate vicinity of a switched segment.

The single types of flagging listed in the above typology will be presented individually below, with examples taken from the Manchester and Milan data. As will become evident by looking at the examples, more than one type of flag may co-occur with one single switch. In the examples to follow each type will be presented individually, even though it may be accompanied by other flags.

### 5.1.1 Paralinguistic markers

Paralinguistic markers include non-linguistic verbal phenomena such as laughter, nervous coughing or other signals invested with communicative value. Paralinguistic markers tend to follow, rather than precede the switch, as in the

example from the Manchester data reproduced below in (20), in which the inserted adjective *biased* is followed by speaker's laughter.

20. \*NIC: il mio esempio è biased@en &=laughs .  
DET my.M.SG example be-PRS.3SG

*'my example is biased'*

[Manchester 4]

There are cases, however, in which the paralinguistic markers precedes the switch, such as (21) below, in which an audible sigh immediately precedes the switch to the Italian adjective *difficile* 'difficult'.

21. \*JON: learning # Italian from learning English it's ## &=sighs difficile@it .  
difficult

[Milano 6]

### 5.1.2 Unfilled pauses

An unfilled pause (UP) is simply a moment of silence within an utterance. While some interruptions of speech delivery are physiological – for instance to allow for breathing – others may represent a form of hesitation. When looking at data, it may not always be possible to distinguish between the two, unless the pause occurs at a clearly identifiable prosodic boundary. Unfilled pauses are classified here based on their length as short, medium or long. The length of the pause has been transcribed impressionistically in the present work, as discussed in section 5.5 below. Unfilled (or silent) pauses are marked by one or several # symbols, as shown in (22).

22. \*EMM: they want # bums on seats and they want live # [/] <live bodies in # aula@it> [>].  
classroom

[Milano 6]

The Italian word *aula* 'classroom' is preceded by an unfilled pause – marked by the symbol "#". Since the pause does not occur at a phrase or turn boundary, it is interpreted as a form of flagging or hesitation.





### 5.1.6 Explicit pre-empting (pre switch only)

In some instances speakers explicitly warn the interlocutor that they are encountering a lexical retrieval problem and that they are searching for a word. Sometimes the sequence may be interpreted as a request for collaboration to complete the utterance, and that the interlocutor may or may not respond. In other cases, the sequence seems to have the function to avoid a switch (as noted in Muysken, 2000), although this still occurs at the end of the sequence. The example in (27) illustrates this occurrence.

27. \*ANT: o ci sarà una: uhm@0 # una &k co(me)  
 or there be-FUT.3SG DET IM DET how  
 come si [//] dice una: # partial@en completion@en<sup>21</sup> certificate@en [>].  
 how IMP says DET partial completion certificate

*'or there will be a uhm a how how do you say a partial completion [sic] certificate.'*

[Manchester 1]

In the example above, the speaker engages into a rather complex flagging sequence while looking for a word. The sequence includes lengthening, retracing, UPs and NLFPs. Of interest here, however, is the explicit *come si dice* 'how do you say', through which the speaker signals his lexical search – and possibly sends a request for help to his interlocutor.

### 5.1.7 Explicit justification (post switch only)

Justification represents in a sense the mirror image of pre-empting, in that it also constitutes an explicit commentary on a switch. Unlike the previous type of flagging, which seems to be a way to avoid switching, justification represents an account of the switch or a qualification of the switch that has just occurred as a marked or non-optimal lexical choice. An example is given in (28) below, in which the speaker qualifies his switch *co-advisor* by following it with *o qualcosa del genere* 'or something like that', implying his lexical choice may have been approximative.

---

<sup>21</sup> The speaker says *completion* [sic] rather than *completion*.

28. \*GIO: ah@0 e quindi Sam@0 Moore@0 ## < è il terzo:  
 IM and so be-PRS.3SG DET third  
 # co-advisor@en o qualcosa del genere> [>] .  
 co-advisor or something of.DET kind

*'ah and so Sam Moore is the third co-advisor or something like that.'*

[Manchester 4]

In other cases, justification may be more explicit and be accompanied by a translation equivalent.

### 5.1.8 Disfluencies, false starts, retracing

This phenomenon is known with several labels in different sub-fields within linguistics. Pragmaticians use the notion of 'recycling' (see Rieger, 2003; Fox *et al.*, 2010), while in other areas the idea of 'maze' (see Fehringner and Fry, 2007) is sometimes used. Here we will mostly refer to the phenomena as retracing or false starts. This includes any interruption of speech delivery followed by the repetition of a phrase or utterance, either identical or with changes. The switch is normally found at the end of the retracing sequence, when the normal delivery of speech effectively resumes. An instance of the phenomenon is shown in example (29). Retracing is marked by the symbol [/] when the retraced segment is repeated *verbatim*, and by [//] and [///] when the retraced segment shows minor or major changes respectively. Angle brackets mark the scope of the retracing.

29. \*KEL: <he's such a> [<] [//] # he's <just a> [//] <such a> little cucciolo@it .  
 puppy

*'he's such a, he's just a, such a little puppy'.*

[Milano 2]

### 5.1.9 Vowel/syllable lengthening

Vowel and syllable lengthening are sometimes also described as draws (Fehringner and Fry, 2007). As the name implies, they constitute the elongation of a vowel within a syllable or of the entire syllable, usually of the word or

hesitation marker immediately preceding or following the switch. They are sometimes accompanied by a decrease in loudness (trailing off). Lengthening seems quite common in the L1 Italian data (Manchester), while it appears less frequently in the English L1 data (Milan). The example in (30) illustrates vowel lengthening on the Italian feminine singular determiner *una* ‘a’, before a switch. Lengthening is marked by a colon symbol “:” after the lengthened vowel or syllable. Note how the switch is accompanied by retracing of the determiner as well.

30. \*MAR: <lui ha [ɪ] ha mandato una:> [<] una mass@en  
 he have-PRS.3SG have-PRS.3SG send-PST.PTCP DET DET mass  
 mail@en chiedendo ### uh:@0 insomma questo incontro .  
 mail ask-GER IM in.short this meeting

*‘he has has sent a a mass mail asking for uh well this meeting.’*

[Manchester 3]

## 5.2 Data transcription

The transcription of the data was carried out using the CHAT<sup>22</sup> system (see section 5.3). At the time of writing, over half of the data has been transcribed in full. **Manchester 1**, **Manchester 2** and a section of **Manchester 4** have been transcribed entirely. **Milano 2**, **Milano 3**, **Milano 4** and **Milano 5** have also been completely transcribed. For the rest of the data, due to time constraints, only the bilingual sections have been transcribed at the time of writing, so that the analysis could be carried out. Samples of the transcripts are presented in **Appendices 8** and **9**. The initial focus of this study was on more grammatical aspects of CS. This is partly reflected in the transcription conventions adopted, such as the choice of the clause as the unit of transcription, as detailed in section 5.3.1. However, the project had been devised to look at more than one theoretical framework from the outset, meaning that the way in which the data have been transcribed is still adequate for the study of flagging, both from a quantitative and a qualitative standpoint. Theoretical and practical guidelines used in transcription, as well as details of the information encoded in the transcript, are discussed in the following sections.

---

<sup>22</sup> <http://childes.psy.cmu.edu/>

### 5.2.1 General principles for transcription of the data

Transcription of verbal data is never a theory-neutral enterprise (see Ochs, 1979). A transcript is not simply the written transposition of a recorded interaction, since it is virtually impossible to encode all the features of speech in writing. Transcription is by its very nature selective, i.e. it encodes and shows only those features of the discourse which are relevant to the analyst's research agenda. In other words, transcription is always informed by a theoretical framework. A transcript could potentially include a great deal of information about the data it represents; besides making the transcription process more time-consuming, however, a transcript which is burdened by too much information would not be easily accessible and not very useful. It is therefore preferable to select the traits that are going to be encoded in a way that is consistent with the aims of one's work. The conventions adopted for transcribing the data are largely based on the criteria originally developed for the Welsh-English Bangor Siarad Corpus, which is publicly available online<sup>23</sup>. Some adaptations were necessary in order to make the data suitable for the study of flagging from both a quantitative and a more qualitative (namely CA) perspective. The adaptation of the existing conventions followed three general criteria: comparability with existing data, interdisciplinarity and time effectiveness. Inevitably, these three criteria sometimes conflicted with one another. As far as comparability with existing data is concerned, the choices made here were largely informed by work carried out at the ESRC Centre for Research on Bilingualism at Bangor University<sup>24</sup>. Several bilingual corpora involving different language pairs have been collected in recent years by the Centre in order to test competing theories of CS. Keeping the coding of the present data close to existing corpora was thought to make comparability between data sets easier in future research.

Comparability as a guideline was at times directly in conflict with the proposed aim of the present project to look at the data using both quantitative and qualitative approaches. The Bangor conventions have been designed chiefly for work on grammatical aspects of CS, while in the present work greater attention

---

<sup>23</sup> <http://www.talkbank.org>

<sup>24</sup> <http://www.bilingualism.bangor.ac.uk>



is given to more eminently discourse phenomena, such as hesitation markers, overlaps and lengthening of syllables. The increased degree of attention given to these features, however, does not reach the level of detail achieved in work carried out within the Conversation Analysis tradition. The system adopted here may be seen as an intermediate solution, which allows both a formal and a conversational analysis to be carried out.

The increased degree of attention to discourse phenomena makes the transcription inevitably more time-consuming than a system focussed only on more grammatical aspects of CS. The alignment of overlaps in particular has a definite impact on the amount of data that can be transcribed in a given time period. Since a relatively large amount of data had to be coded and only limited resources were available, it was necessary to limit to an extent the amount of discourse information encoded in the transcript. At the same time the transcription of bilingual passages was prioritised.

The balance between the three constraints results in a composite transcription system: while the choices adopted were sometimes pragmatically motivated, special attention was given to ensure that the final transcript is as much as possible an internally coherent system, and that it is also accessible to scholars with different research agendas.

### **5.2.2 Transcription system**

After sessions with participants, files were transferred from the digital recorder onto an external hard drive. The recordings were made in .mp3 format and were subsequently converted into .wav format. Brief notes were also made at this stage about each recording, its exact setting and any information about participants that was deemed to be useful.

Each file was subjected to some basic editing. Initial and final parts were removed – these cuts were limited simply to sections in which the researcher was still present and talking to participants or greeting them. Participants were given the option of editing out sections of the recording if they wished, but none of the participants requested this. The volume of the recordings was normalised and

noise reduction filters applied when necessary. In some recordings the different position of microphones meant that one voice would be considerably louder than the other. This was also addressed whenever possible.

Following the editing and standardising of the audio files, recordings were ready for transcription. Transcription was carried out in the CHAT format, the coding standard adopted by the CHILDES<sup>25</sup> project – Child Language Data Exchange system (MacWhinney, 2000). As the name suggests, the CHILDES enterprise was originally developed and is still employed for the analysing and sharing of child language acquisition data, and was not explicitly devised for the study of multilingual interaction – although a section of CHILDES deals explicitly with child bilingual acquisition. All the data that is part of the CHILDES project has been made available online.

CHILDES is now in turn part of the larger project Talkbank<sup>26</sup>, also maintained by MacWhinney, whose aim is to provide an interdisciplinary system for studying and sharing conversational interactions in both children and adults.

The CHAT system was first adapted to adult bilingual and multilingual data by the LIPPS Project (Language Interaction in Plurilingual and Plurilectal Speakers). The LIPPS project was also the propelling force behind the LIDES (Language Interaction Data Exchange System). A LIDES manual was published as a special issue of the *International Journal of Bilingualism* (2000 :4 :2). The manual details the rationale for adopting the CHAT coding system and provides a step-by-step guide to transcribing spontaneous bilingual data. A more concise and updated version of the manual is to be found in Gardner-Chloros (2009).

The CHAT system was designed to be flexible and accommodate the needs of researchers with different agendas. Provisions are made, for instance, for researchers wanting to adopt CA conventions, or for the inclusion of gestural and proxemic information. The LIDES conventions have a similarly wide scope, by also allowing for the encoding of data involving different theoretical orientations and research agendas.

---

<sup>25</sup> <http://childes.psy.cmu.edu/>

<sup>26</sup> <http://talkbank.org/>

### 5.3 Features of the CHAT system and transcribing conventions

The main advantage of transcribing data in the CHAT system rather than simply using a conventional word processor is that the transcript is completely machine-readable. By using the CLAN software, which is freely available as part of the CHILDES project, analysts can run searches on one transcript at a time and also entire corpora, simply by entering strings of commands into the program.

This allows analysts to automatically extract all instances of bilingual utterances from the data, or to discover the frequency of specified items or categories of items. Needless to say, the range of options available depend heavily on the amount of information that has been included in the transcription.

A further feature of the CHAT system is the possibility of linking the transcript directly to segments of the original audio file. This is done through a system of bullets which are inserted manually by the transcriber and make it possible to navigate a transcription and listen to a certain portion of the recording starting from any point. Continuous playback is also available, which the user of the transcript can employ to follow the unfolding of the interaction very closely.

CHAT transcripts have a standardised format, characterised by the presence of several compulsory initial headers. The headers contain information about the languages used, the participants, the general situation in which the recording took place, the date and duration of the file and details about the transcriber(s). A series of optional fields can also be included. In the present study, comment fields were added to the header containing details of the language tagging conventions adopted. These are discussed in detail in section **5.3.2**.

CHAT transcripts are also characterised by the possibility of adding so-called dependent tiers. Each line of speech, headed by the speakers' name, constitutes the main tier. To this main tier it is possible to add dependent tiers. These are placed immediately below the main tier and contain additional information about the utterances, such as addressee specification, word-by-word morphological glosses, translations, gestural and proxemic information or contextual comments. Use of dependent tiers was quite limited for the present data, with the exception of comment tiers. Comment tiers were typically employed when speakers used non-standard forms, to signal the presence of external participants or add any

accessory information that could facilitate access to the transcript. An example is given in (31) below; a comment (shown in bold) is inserted in the transcript to explain the meaning of a term which may be unfamiliar to most readers.

The comment tier refers to the immediately preceding speaker tier.

31. \*KEL: I'm trying to think of some games .  
\*KEL: oh@0 if you want when I come back Mad\_Libs@0 are really fun .  
%com: **"Mad Libs" are a kind of word game for children.**  
\*KEL: xx do you know what they are ?  
\*ANN: no .

[Milano 4]

### 5.3.1 The body of the transcript

As already stated in section 5.2.1, the present corpus was modelled on work already carried out by the corpus-based group at the ESRC Centre for Research on Bilingualism at Bangor University. The conventions of the Bangor group were followed in selecting the clause as the unit of transcription. This allowed comparability with the other Bangor corpora. Even though the focus of this project later changed to a less grammatical and more discourse-oriented approach, the format of the transcription was not changed.

While some of the dynamics of interaction may be lost when transcribing clause-by-clause, these features are for the most part not relevant when looking at flagging, which happens in the immediate environment of a switched segment (typically within a clause rather than at clausal or turn boundaries). Moreover, while the turn was not used as a unit of transcription, conversational turns are still identifiable in the transcript as groups of adjacent clauses.

These considerations notwithstanding it should be observed that the choice of the clause caused some difficulties; spontaneous speech is not always readily divisible into discrete clauses, which means that at times, arbitrary decisions had to be made. Furthermore, in some instances, speakers used phrases instead of full-clauses, in which case, phrases and fragments were used as lines of transcription.

In transcribing clause-by-clause, subordinate clauses are kept on the same line as the main clause that they depend on. Coordinate clauses, on the other hand, start a new line. Extra-clausal or peripheral elements, such as paralinguistic or discourse markers, are normally kept on the same line as the clause they immediately precede or follow. Each new line is marked by a three-letter abbreviation of the speakers' name, and is repeated at each new clause, even within the same speech turn. A screenshot of a CHAT window is reproduced below.

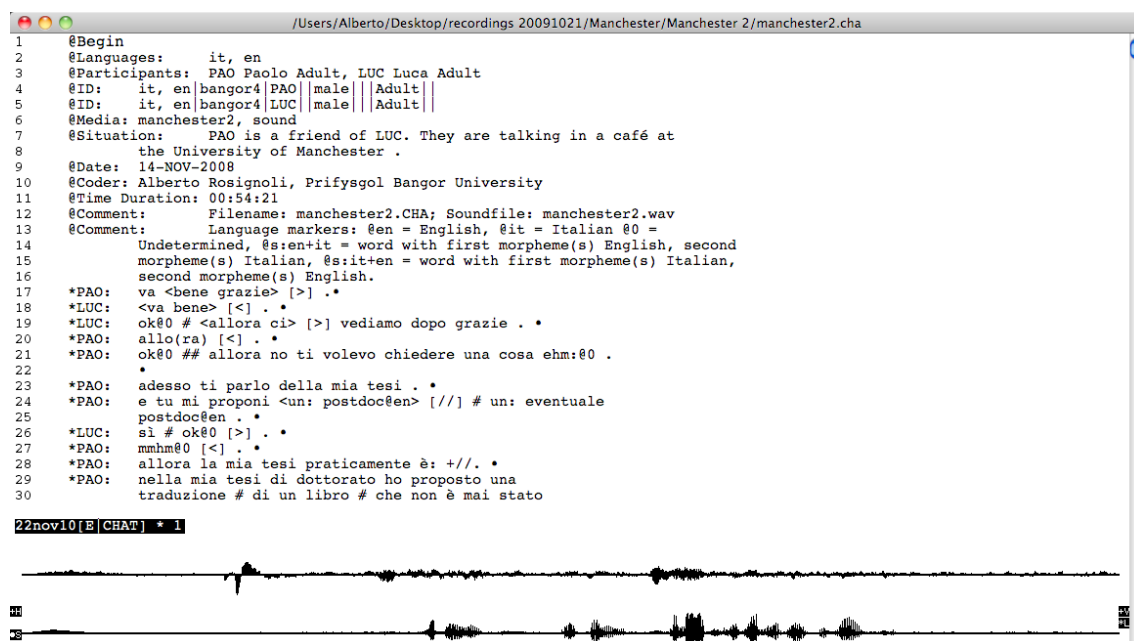


Figure 1. Screenshot of a CHAT transcript with file header and soundwave.

### 5.3.2 Language tagging and ambiguous items

One of the features of the LIPPS and LIDES project which is supported by the CHAT system is the possibility of tagging each word for language membership. According to the original guidelines of LIDES, an “-@” symbol is attached directly to the end of each word, followed by a number for each language. If the number 2 is assigned to English and 4 to Italian we would have, for example, *word@2* for English and *parola@4* for Italian. These conventions are illustrated in Gardner-Chloros (2009).

Besides a category for each language, an extra category is added to mark all those items whose status is ambiguous. The symbol used in Gardner-

Chloros (2009) is “@0”. This category includes proper nouns, brands, names of organisations and of works of intellect (films, books etc.) when a translation is not available in the other language. Place names are marked as ambiguous when there is no other-language version of the name. For example, the proper noun *California* is always marked as ambiguous in English-Italian speech even though a difference can sometimes be heard between a native English and a more Italian pronunciation of the word. *Londra* ‘London’ and *London*, on the other hand, would be marked as Italian and English respectively. Interactional markers, such as NLFPS *uhm* or hesitation markers are also marked as ambiguous when they are present in both languages, in spite of their exact phonetic realisation being sometimes different in the two languages. This is the case of *oh*, which is used in a similar range of domains in both English and Italian. The marker is normally realised as a falling diphthong in English – most typically either [əʊ] or [oʊ], although a monophthong [o] is also possible in some varieties. In Italian, it is consistently realised as a monophthong, normally [o]. Even when a difference can be perceived, *oh* is consistently tagged as an ambiguous item.

On the other hand, when interactional markers are only present in one of the two languages, as in the case of English *ouch*, the tagging reflects their specificity. In a recent update of the CHAT system, a default language system has been included. This new system has not yet been widely adopted in the set of bilingual data part of LIDES, but is being progressively implemented in new transcriptions. In the new system each transcript is identified as having a principal language (to be understood simply as the language which is likely to be used most frequently). This language is declared in the header file and is not marked on every word, unlike in the original system. The less-frequent language, on the other hand, is marked with “@” followed by a two-letter code for the language – in the present case “@it” for Italian in transcripts with English as a main language and “@en” for English in transcripts with Italian as the main language. The tag for ambiguous markers is “@s:en&it”, although the pre-existent “@0” has been maintained here as a more legible alternative. In the case of words containing morphemes from both languages, a system of composite markers is used. For instance, a word whose first morpheme is Italian and the second English

is marked as “@s:it+en”, with the order of language codes reflecting the order of morphemes. Such examples, however, are limited to very few occurrences<sup>27</sup>. The new system proves to be advantageous in terms of readability and time-effectiveness, since the transcriber does not have to tag every single word in both languages. From a theoretical standpoint, the identification of a “default” language may be criticised, on the grounds that the analyst is expected to establish *a priori* which language is used more-frequently, normally based on a sample of the conversation. Further, it is possible that some bilingual conversations have roughly a 50/50 split between the two languages, therefore undermining the choice of an operational default. For the present data, however, this system proved to be particularly useful: all conversations are conducted for the most part in the speakers’ L1, with the most frequent switches being one word insertions or occasional short sequences. The identification of a default language is therefore particularly suited to the data, since the preponderance of one of the two languages is the norm.

These methodological considerations, however, underlie a strictly theoretical question: how is it possible for the analyst to ascribe words to either language or to both? In the present work, a dictionary criterion was adopted. Any words whose status is ambiguous would be searched in an Italian and an English monolingual dictionary. If a word appeared in both dictionaries, it was marked as ambiguous. If, on the other hand, it only appeared in one dictionary it was assigned to the corresponding language. The dictionaries used were *Il Sabatini Coletti – Dizionario della Lingua Italiana* for Italian and the *New Oxford American Dictionary* for English. Both dictionaries can be consulted electronically.

This method of establishing language membership potentially presents a series of problems: recent words that have entered the lexicon of a language through borrowing may have not entered the dictionary simply due to the fact that the compilation and updating of dictionaries offers a delayed image of the lexicon of the language. More importantly, however, some entries may be included

---

**27** On some occasions, speakers also used words or phrases from languages other than English or Italian. These words have been tagged with as instances of non-default language using the appropriate code for each language (e.g. “@sp” for Spanish).

or excluded depending on the orientation of the lexicographers compiling the dictionary, which could be more or less purist. This issue is particularly relevant with reference to English loanwords entering the Italian lexicon. Due to the influence of Anglo-American culture starting in the second half of the twentieth century, thousands of English-origin words have entered the Italian lexicon. These borrowings belong to very different domains, including but not limited to business, technology, politics, sport and entertainment (Antonelli, 2007). Increased widespread knowledge of English as an L2 in Italy may be playing a part in this phenomenon, but it should be observed that many of these borrowings are also routinely used by monolingual speakers of Italian. A sample from the Manchester data includes words such as *escalation*, *vamp*, *online*, *tutor*, *input*, *horror*, *skinhead*, *fiction*, *network*, all of which are attested in the Italian dictionary. Perhaps unsurprisingly, most of these borrowings are nouns. This extensive borrowing process is still very much underway, with new items continuously entering the language at all levels, from institutional usage to professional jargons and youth culture. It could be argued, therefore, that the situation being very complex, the dictionary criterion may present some limitations when one is trying to ascribe language membership to a given word. A case in point is the word *plot*, which appears once in the data and is attested in the Italian dictionary. *Plot* has a rather specialised meaning in Italian (it is only used to mean 'storyline' in a film) and may only be known and employed by Italian speakers as part of a professional jargon. It could be disputed whether the word would be known or used by an average speaker of Italian, who may only know Italian equivalents *trama* or *sceneggiatura*. For reasons of consistency, however, all words listed in the dictionary were treated as ambiguous words, regardless of their perceived currency with monolingual speakers of Italian. The reverse of this situation, i.e. the case of Italian loanwords into the English lexicon, is not quite as problematic. Italian borrowings into English are rather circumscribed in domain, and their status as part of the recipient language is quite undisputed. There is a third category of loanwords which are treated as ambiguous, i.e. words from languages other than English or Italian which have been borrowed into



both languages and are found both in English and Italian dictionaries. That is the case of words such as Japanese *shiatsu* or Portuguese *samba*, both of which are encountered in the corpus.

### 5.3.2.1 The use of phonological criteria in ambiguous cases

In some instances, it would have been possible to tag ambiguous words as belonging to just one of the languages based on their phonetic or phonological realisation. This would be the case with established borrowings such as *computer* or *Internet*, which are realised sufficiently differently in the two languages to be classified as either English or Italian words. Nevertheless, it was decided not to rely on phonological cues for borrowings when assigning a language tag. First of all, unlike presence in a dictionary, phonological adaptation is not a binary category. Ambiguous words may present different degrees of phonological adaptation into the host language and fall into a grey area, with some segments approximating source language characteristics, and others being closer to those of the recipient language. Finally, the speakers in this study are almost all L2 speakers of either English or Italian; it is therefore possible that they did not possess a native-like phonological system for their L2. This means that their L2 productions could always exhibit the interference of L1 phonology. Even when asked, not all speakers may always be able to produce separate “Italian” and “English” versions of ambiguous words. One example would be the case of a word such as *computer*. Speakers of British English and American English produce a form close to [k<sup>h</sup>əm'pjʊ:tə] or [k<sup>h</sup>əm'pjʊ:tə] respectively. Italian speakers, on the other hand, would typically realise the word as [kom'pjuter(ə)], with no aspiration of the initial [k], replacement of the unstressed central vowels with Italian [o] and [e] due to the influence of orthography and final alveolar trill. The English and Italian versions are theoretically different enough for the analyst to distinguish between the two; in practice, however, speakers sometimes produced intermediate versions, making the identification of the language very difficult. For these reasons, even in those cases in which the phonological cues pointed to one of the languages, it was decided to disregard these cues and follow a more

conservative tagging system, with all ambiguous words being treated as belonging to both languages. The only exception to this rule is the case of homographs which are not English-Italian contact forms (like *computer*) and which have sufficiently distinct realisation in the two languages, such as *zebra*: English [ˈzebɹə] or [ˈzi:bɹə], Italian [ˈdzebra] or [ˈdzebra]. In this case, phonological criteria were adopted to mark the language, since the words were not perceived as ambiguous.

### 5.3.3 Interactional markers

Interactional markers have been treated and tagged like other words whenever possible. Some interactional markers are highly conventionalised and have standard written forms and dictionary entries. Such is the case for *ah* and *oh*, which are found in both languages and used with a similar range of functions. Italian-only examples include *ehm* [ɛm] and *ahi*, [ai] which are also found in a dictionary. Other interactional markers lack a dictionary entry, although sometimes there is a somewhat conventional spelling for them which is used when transcribing spontaneous speech, for instance in works of fiction. These spellings have been retained as much as possible in the transcript in order to make the data more accessible to speakers of the two languages.

For all other interactional markers, the lists presented in the CHAT manual have been used as much as possible, so as not to introduce too many new elements or spelling variations into existing conventions. Interactional markers lacking dictionary entries but which are not considered to be language-specific, such as the agreement marker *mhm* [m.hm] are always treated as ambiguous.

### 5.3.4 Pseudonyms

All participants are assigned a pseudonym by the researcher to ensure their anonymity. Further, all other individuals mentioned in the recording are also assigned a pseudonym, with the exception of public figures, celebrities, fictional characters and the researcher's name. When assigning pseudonyms, the gender

and language identity of the original name are preserved: speakers with Italian names have been assigned Italian pseudonyms only, while speakers with English names have been assigned English pseudonyms. In those instances in which someone's full name is mentioned, a full name-plus-surname pseudonym is inserted, always ensuring anonymity. A further requirement in the choice of pseudonyms is the fact that in the CHAT system each speaker is identified by a three-letter abbreviation – normally the first three letters of their pseudonym. Pseudonyms that differed in the three initial letters were therefore chosen in order to avoid confusing speakers' identities.

Inter-transcription consistency in the assignation of pseudonym has also been maintained whenever possible. When the same person is mentioned in more than one recording, the same pseudonym is used across all the transcripts in which they appear. This is done to provide a more realistic representation of the data, as well as preserving important information about the speakers' social network.

### 5.3.5 **Other transcription conventions**

In this section, some additional conventions used to transcribe the present data are briefly presented.

#### 5.3.5.1 **Pauses**

In transcriptions following CA conventions, pauses are typically marked with a great degree of precision and their duration measured in tenths of seconds. A more impressionistic system was adopted here, since CA is not the sole analytic framework used. Pauses are marked with the hash symbol “#”, and the repetition of the symbol indicates increased duration. A micro-pause is marked with a single hash “#”, and longer pauses are marked with two “##” or three “###” hashes, depending on their perceived length. Variation in the speech rates of different speakers may result in some inconsistencies in the marking of pauses: pauses in speech that is generally perceived as slow may be treated as shorter than a pause of the same duration by a speaker with a higher speech rate.

Pauses between utterances are not marked, with the exception of prolonged periods of silence between the speakers, which are noted with a separate entry in a comment line in the body of the transcript.

### 5.3.5.2 Spoken and colloquial forms

Transcription of speech is orthographic, with little or no phonetic or phonological detail marked. Most non-standard pronunciations are not marked orthographically, even when there are established conventions for signalling them – as is the case [h] dropping in English. Partial exceptions to this rule are detailed below. In English, verbal contractions of the type *gonna*, *wanna*, *gotta* for “going to”, “want to”, “got to” are transcribed as contracted when heard, but are always followed by an uncontracted gloss, as in the example in (32):

32. \*ANN: and how are you **gonna** [: going to] survive ?

[Milano 4]

Instances of optional apheresis and apocope (i.e. deletion of initial or final segments or syllables) in both languages are marked by including the elided segments in round brackets, as shown in the examples below:

33. \*AMI: (be)cause we had an individual # meeting .

[Milano 3]

34. \*NIC: Allora chi è                    Giovanna@0 #    che    conosce:  
           then    who    be-PRS.3SG    Giovanna            REL.PRN    KNOW-PRS.3SG  
           (que)sto        ragazzo ?  
           this            boy

*‘so who is this Giovanna who knows this guy?’*

[Manchester 4]

Provided that the target word is discernible, the same notation is also employed when a word is left incomplete because the speaker has been interrupted, is trailing off or is retracing his or her speech. When incomplete words are not discernible, a broad phonetic transcription of the fragment is used instead.

Cases of obligatory or systematic final vowel elision in Italian are marked using standard orthographic conventions (i.e. apostrophes) as in the example below, in which the final vowel from Italian adjective *bello* ‘beautiful’ is elided.

35. \*SAR: <ha trovato un bell’ ambiente> [<] caldo  
 have-PRS.3SG find-PST.PTCP DET nice environment warm

*‘(she) has found a nice and warm environment.’*

[Manchester 1]

A slightly different case is that of the Italian optional removal of final unstressed vowels with some verbs forms, both finite and non-finite. The precise rules of this omission are rather complex, and a certain degree of variation is found amongst speakers of different varieties of Italian (see Bertinetto and Loporcaro, 2005). These forms, however, are considered part of the standard language. They have been transcribed with no special marking, since that is how they are normally encountered in standard written Italian. An example is given below, where the verb *essere* ‘to be’ is transcribed as *esser*, with the final vowel removed.

36. \*LUC: s̀ì [<] quindi dovrebbe esser pi`u` tedesco dei tedeschi .  
 yes so must-PRS.CONDIT.3SG be-INF more German-M.SG of.DET.PL German-M.PL

*‘yes, so he ought to be more German than the Germans’*

[Manchester 2]

### 5.3.5.3 Non-standard and regional words

On some rare occasions, speakers use words from regional Italo-Romance languages, which are not found in the Italian dictionary. One example is the Neapolitan noun *quaglio* ‘boy/lad’ [Manchester 2], cf. standard Italian *ragazzo*. For practical reasons these words have been tagged as Italian words. While they are not always found in the dictionary there is no ambiguity with regards to their language source: including them in the ambiguous category would not be accurate. Besides, many of these words are routinely used in regional varieties of Italian. The occurrence of words from regional languages is noted by the insertion of a comment tier below the main tier.

#### 5.3.5.4 Overlapping speech

Overlaps are a frequent feature of spontaneous conversation. The CHAT system offers various alternative ways of marking overlaps in a more precise or broad way. Like most aspects of the transcription, the choices made when marking overlaps depend greatly on the investigators' research goals and theoretical orientation. In the present work, an intermediate system has been adopted between a narrow and a broad notation. Whether for single or multi-word overlapping segments, the beginning of the overlap is approximated at the level of the word, i.e. there are no intra-word markers of overlap.

The symbols [>] and [<] are used to mark the beginning and end of overlapping speech. In the example below, the word *mishmash* is uttered in overlap by the two speakers. The [>] symbol in JON's utterances indicates that the preceding word is overlapping, while [<] in EMI's line indicates the end of the overlap.

37. \*JON: but ## there are times where it just becomes this # mishmash [>] of everything .  
\*EMI: mishmash [<] .

[Milano 5]

When the overlap lasts more than one word, angle brackets are used to indicate the scope of the overlap. In the case of multiple overlapping segments within a single utterance, the overlap markers are indexed with progressing numbers [>1], [>2], so that it is possible to distinguish the different points of overlap. This is shown in the example below. The segment <point is [/] is that> overlaps with Anna's chuckling in the second line, while *connected* is in overlap with *she's* in the third line.

38. \*KEL: but # the <point is [/] is that> [>1] <she's uh:@0> [/] # she's well connected [>2].  
\*ANN: &=chuckles [<1] .  
\*ANN: she's [<2] got her feet in the right place [>] .  
\*KEL: she's [<] definitely got her feet in the right place .

[Milano 4]

### 5.3.5.5 Unintelligible words/sections

A question mark symbol in square brackets “[?]” following a word or phrase is used whenever the transcriber is unsure about his transcription. As in the case of overlaps, angle brackets are used to indicate the scope of the uncertain transcription. If a section is completely unintelligible a “xx” symbol is used when the section is likely to be a single word, whereas “xxx” is employed for segments which are thought of as being made of more than one word. In one or two rare cases, longer unintelligible sections of the recording lasting several utterances are found. This is normally due to excessive background noise or other disruptive factors. These sections are marked with a comment line in the body of the transcript.

When a speech segment is not recognised as a word, but its phonetic form is clear, a broad IPA transcription is used. These forms are preceded by the “&” symbol to mark their special status as non-orthographic signs.

### 5.3.5.6 Vowel and syllable lengthening

Lengthening of vowels or syllables (also referred to elsewhere as drawing) is marked throughout the transcript. A colon “:” symbol is used after the lengthened segments as in the example below:

**39.** \*KEL:    **and:** uh@0 her accent is **so:** funny .

[Milano 4]

The use of the lengthening marker is not absolute, i.e. is not based on the objective measure of the segment’s or syllable’s length compared with an average or standard measure. Rather, the marker signals a form of hesitation and a slowing of the speaker’s perceived average speech rate.

## 5.4 Concluding remarks

This chapter has introduced the methods adopted to classify and transcribe the two English-Italian CS data sets presented in chapter 4. First, the operational definition of flagging used in the present work has been presented, together with examples of single types of flagging (adapted from Hlavac, 2006) taken from both the Manchester and Milan data.

The transcription of the data was also illustrated. The methods have been informed by three criteria: comparability with previous work, interdisciplinarity (intended as the use of more than one theoretical framework) and time effectiveness.

The problem of ascribing a language tag to the data has been discussed.

The use of the dictionary criterion is illustrated, together with a review of possible alternatives (i.e. the use of phonological criteria).

The final part of the chapter details the conventions and practical problems faced during transcription which required specific decisions.

It has been stressed how transcription is a necessarily selective and interpretive operation and how the multidisciplinary focus of the present work involved operating some compromise choices when coding the data. These have been made in order to allow the data to be analysed from both a structural perspective and a more qualitative, interpretive perspective, as well as making the data sets accessible to scholars coming from different theoretical backgrounds. ¶



# DESCRIPTION OF SWITCHES AND QUANTITATIVE ANALYSIS OF FLAGGING

In the present chapter, the flagging patterns for switches in the two data sets are presented and analysed. First, various types of switches in the data are described; insertional switches are found to be the prevalent type in the data. Subsequently, switches in the two groups and their flagging rates are analysed in detail, according to their grammatical category. Single nouns, multiword noun phrases, adjectives and verbs are analysed separately. Special attention is given to variation in the rates and type of flagging for each grammatical category both within and between the Manchester and Milan group. Lone nouns are analysed more in detail, and a comparison is drawn with nouns in monolingual contexts.

## 6.1 Method of analysis

The analysis started by extracting all the switches from the transcribed data. The data included both fully transcribed conversations and bilingual fragments transcribed from the rest of the corpus. Since conversations in the Manchester data are overwhelmingly in Italian and conversations in Milan overwhelmingly in English, the directionality of switches shows a high degree of uniformity, with switches going from the speakers' L1 to their L2. The task simply consisted of the extraction of items marked as other-language in each corpus, i.e. English for Manchester and Italian for Milan. Since all of the data had been tagged for language during the transcription phase, extraction of all the examples from the data was a relatively simple and rapid procedure. It is worth remembering here

that for the purposes of language tagging a dictionary criterion was used (see section 5.3.2). Words of English origin found in an otherwise Italian environment would be marked as English only if they could not be found in a dictionary of Italian. When words were found in both Italian and English monolingual dictionaries, the items would receive a ‘zero’ marking, signalling that they could not be exclusively attributed to either language. Instances of words falling in this category are mostly well-established borrowings, such as *email*, *internet*, *party*, in Manchester or *pasta*, *pizza* in Milan. In the case of cognates, phonological criteria were adopted where sufficient distance between English and Italian forms existed. While Italian borrowings into English are a rather circumscribed phenomenon, and are limited to certain lexical fields (food, art, music etc.) the extensive borrowing of English words into contemporary Italian makes the drawing of boundaries not always an easy task. The details of language tagging and the effects for the classification and analysis of the data are discussed in section 5.3.

### 6.1.1 Types of switches in the data

The overwhelming majority of the switched material extracted from both data sets falls within the category of insertions, defined as “something akin to borrowing: the insertion of an alien lexical or phrasal category into a given structure.” (Muysken, 2000: 3; see also section 2.3.1.3). Examples (40) and (41) illustrate this pattern:

40. \*PAO: sì # quella era # più educated@en.  
 yes that-F be-IMF.3SG more educated

*‘yes, she was more educated.’*

[Manchester 2]

41. \*AMI: the vicino@it downstairs will break it uh@0 ah@0 [>].  
 neighbour

[Milano 3]

As implied in the definition of insertion, in both examples the other-language elements (an adjective in (40) and a noun in (41) have simply been slotted (or “nested” to use Muysken’s words) into an otherwise monolingual clause.

While insertions were prevalent, some instances of other types of language-mixing were also encountered, as in example (42):

42. \*KEL: they’re the only people in@0 campeggio@it che@it vanno@it  
in camping.site REL.PRN go-PRS.3PL  
con@it i@it <piedi@it nudi@it # nella@it doccia@it # con@it [>] +/.  
with DET feet bare-M.PL in.DET.SG shower with

*‘they’re the only people in the camping site who walk into the shower barefoot.’*

[Milano 8]

Examples such as the one above are normally referred to in the literature as instances of intraclausal CS or, to use Muysken’s (2000) taxonomy, mixing of the alternational type. In these data at least, alternations seem to constitute a very marginal category compared with the much more frequent insertional switches. In example (42), it is also possible to note that the switching point occurs in conjunction with the preposition *in*, which is a cognate in English and Italian. The Italian preposition has a partially overlapping distribution with its English counterpart – as indeed in the example. It is therefore possible that the switch to Italian was facilitated by the presence of a quasi-homophone cognate, a phenomenon normally referred to as “triggering” (see Clyne, 2003, Broersma, 2009). Instances of triggering, however, are very rare in the present data. Occasionally, speakers also switched from one turn to the next or between distinct utterances within the same turn. In no case, however, did these switches result in extended sequences in the other language. After the turn was completed, the interlocutor would continue in what had been the language used up until the switch, as illustrated in the example (43) below. The speaker Antonio switches to English for an entire utterance, but returns to Italian in his next turn with *perché* ‘why’.

43. \*ANT: <come va il PhD@en> [<] ?  
 \*ANT: gonna@en go@en finish@en it@en ?  
 \*SAR: ah@0 PhD@en è: molto un momento brutto # mercoledì [>] +/.  
 \*ANT: perché [<] ?

- 
- \*ANT: *how is the PhD going?*  
 \*ANT: *gonna go finish it ?*  
 \*SAR: *ah the PhD, it's really a bad moment, on Wednesday*  
 \*ANT: *why?*

[Manchester 1]

## 6.2 Types of insertions

As mentioned earlier, all instances of switches were first extracted from the transcripts, together with the clause in which they appeared. When switches were not part of full clauses, just the line (clause fragment) in which they were found was extracted, as in (44):

44. \*SAR: +” e ancora questi errori di proofreading@en .  
 and again these-M mistakes of

*'and again these proofreading mistakes'*

[Manchester 1]

The extraction produced 225 English switches for the Manchester data over 271 minutes of conversation and 133 Italian switches over 297 minutes for the Milan data. The insertions are mainly one word, and belong to various grammatical categories (lone nouns, adjectives, verbs etc.). The classification of the inserted material into various grammatical categories was not always an easy task, since the boundaries between some grammatical categories present a certain degree of uncertainty in both English and Italian. The most frequent case is perhaps that of past participles, which are often used as verbal adjectives in both languages. When they occur as part of nominal constructions their adjectival status is relatively undisputed. When, on the other hand, they are produced with a copular or auxiliary verb, it may sometimes be not easy to assign them to either grammatical category. Below is an example:

45. \*ANT: ehm: è retired@en # <da due anni> [>] .  
 IM be-PRS.3SG from two years

*'she has been retired for two years'*

[Manchester 1]

The English past participle *retired* is inserted together with the Italian copula – after some hesitation – in an otherwise Italian clause. Ascribing the insertion to either the category of adjective or that of verb in this instance is rather difficult. The verb *essere* ‘to be’ functions in Italian as a copula, an aspectual and a passive voice auxiliary. It can typically be followed by nouns, adjectives and verbal participles; in (45), therefore there are few cues regarding the status of the insertion. The closest Italian translation would not use a verb or an adjective after the verb *essere*, but the prepositional phrase *in pensione* ‘in retirement’. This also does not offer much insight on how the insertion should be classified, as it is not possible to assume that the speaker is using the English insertion to replace an Italian equivalent belonging to the same category. For reasons of consistency, it was decided to classify ambiguous insertions of participles in copular construction of this kind as adjectives.

In some instances, when inserted into the other language, items are used in a different category to the one they normally have in their source language; this is the case when items have undergone grammatical conversion. The process of conversion is productive in both English and Italian, although it targets different grammatical categories. Conversion is particularly common in English (Aitchison, 2003: 177), especially in the case of nouns being converted into verbs.

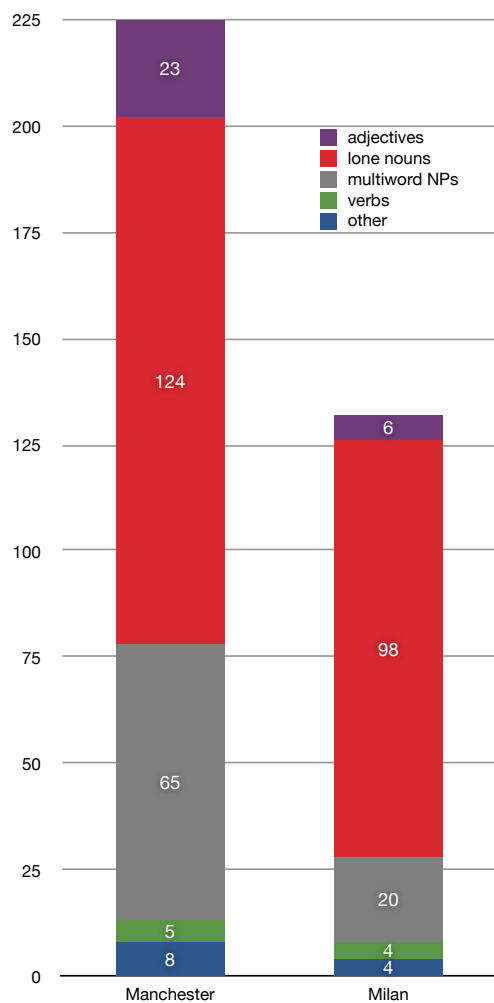
An example is given in (46).

46. \*EMM: <but <she's got to>> [<] [//] # surely she has to studio@it something .  
 study

[Milano 2]

Speakers in the conversation from which the example is extracted have been talking about a *borsa di studio*, lit. ‘bursary of study’. Shortly afterwards, speaker Emma

produces the sentence in (46), in which the Italian noun *studio*<sup>28</sup> ‘study’ is inserted in verb position as part of the English periphrastic construction *to have to + verb*, which expresses obligation. Following English grammar, the noun is converted into a verb with no necessity to add morphology. This is unlike Italian, where verbal inflection would always be expected. The switch *studio* was therefore classed as a verb insertion for the purpose of the analysis, since it is used as a verb in speech. A general picture of the material inserted in the two data sets divided by grammatical category is presented in **Figure 2**.



**Figure 2.** Insertions in the two data sets classified by grammatical category.

---

<sup>28</sup> The Italian noun *studio* is also an established borrowing into English, although not with the meaning employed here. In the present example, it seems highly improbable that the word had been used in its English meaning; it was therefore decided to tag it as an Italian word.

In both data sets nominal constructions constitute the most-frequently inserted morphological category. This includes both lone nouns (also referred to as single nouns<sup>29</sup>) and multiword NPs. Lone nouns make up 56% of the insertions in the Manchester data and 74% in the Milan data. If complex NPs are added to lone nouns, the picture that emerges in both groups is one of very strong preponderance of the insertion of nominal constructions, with 86% of insertions in Manchester and 89% in Milan belonging to this category.

Adjectives are the next most frequently inserted category, although their numbers are much smaller than those of nominal constructions. Both predicative and attributive adjectives constitute 10% of inserted forms in the Manchester data but only 5% in the Milan data, with 23 and 6 tokens respectively. Differences in the placement of attributive adjectives across the two languages may partly explain the smaller number of instances for this category compared with the previous ones. The remaining insertions (4% in Manchester and 6% in Milan) are represented by verbs, discourse markers and interjections and adverbs. Of these, only verb insertions are represented in both corpora, with five English tokens found in Manchester, and four Italian tokens in Milan.

Overall, the picture that emerges is one of relative similarity in terms of elements inserted into either language by speakers in the two groups. The main difference to be noted is that, in Manchester, a higher diversification of inserted elements can be observed.

Before analysing flagging, however, examples of the different types of insertions found in the data will be presented below according to grammatical category.

### 6.2.1 Noun phrases and lone nouns

Lone nouns represent by far the most common category of switched items both in Manchester and in Milan. While as nominal constructions they constitute a subcategory of NPs, they will be analysed separately from the former. This is because they exhibit a different behaviour from multiword NPs, as will become apparent in the analysis.

---

<sup>29</sup> By “single” or “lone” nouns we mean bare nouns, i.e. nouns inserted with no determiner or other modifiers from the same language.

In terms of absolute quantities, single nouns are more frequent in the Manchester data, although they constitute a higher proportion of switches in Milan (see **Figure 2** above). The high frequency of noun insertions is consistent with results from several bilingual corpora in the literature.

Below in (47) and (48) are two examples of single noun insertions:

47. \*PAO: io avevo una ehm # una landlady@en ## che:  
 I have-IMF.1SG DET IM DET landlady REL.PRN  
 era napoletana.  
 be-IMF.3SG Neapolitan-F.SG

*I used to have a uhm a landlady who was Neapolitan'*

[Manchester 2]

48. \*KYL: ah@0 you should have gone to the English nido@it .  
 crèche

[Milano 3]

### 6.2.1.1 Complex noun phrase insertion

After lone nouns, multiword NPs are the second most common type of inserted material found in the corpus. This is true for both data sets. Most inserted NPs are made up of a head noun plus an attributive modifier. Modification by a prepositional phrase is also possible, although less frequent. Determiners hardly ever appear as part of the switched phrase (hence the preference for the label “NP”). When determiners would be expected, they tend to occur in the language of the rest of the clause. As a matter of fact there is only one example of an inserted noun phrase where the determiner is also switched in the Milan data, and no examples in Manchester (if one excludes the occasional quotation of titles of films or books).

A certain number of items in this category is made up of collocations and stock phrases, such as *bad words*, *social sciences*, again consistently with several other existing corpora in the literature. Below are two examples illustrating the prototypical NP insertion, with adjective modification and the switch point after the determiner (Italian *le mie* ‘my’ fem. plur. in (49) and English *these* in (50)).



49. \*SAR:    porto            le        mie        dancing@en    shoes@en    però .  
               take-PRS.1SG    DET       my-F.PL       dancing       shoes       but

*'But then I'll take my dancing shoes'*

[Manchester 1]

50. \*AMI: yeah so the idea # a few years ago was that they # introduced these: #

laboratori@it    facoltativi@it .  
 laboratories     optional-M.PL

[Milano 3]

In both NPs in (49) and (50) we see more than the juxtaposition of two words inserted from the other language, in that the NPs show internal structure not complying with that of their recipient language. *Dancing shoes* and *laboratori facoltativi* in fact, obey the requirements of their source language with regards to head-modifier order. In (49) the NP is pre-modified, consistently with the grammar of English, while in (50) post-modification is found, as would be expected in Italian. Morphological marking also comes from the source language (English plural marking *-s* in *shoes* and Italian masculine plural marking *-i* in *laboratori facoltativi*). Within the framework of the Matrix Language Frame Model, these insertions would be classified as Embedded Language Islands (Myers-Scotton, 2002: 139), in that they exhibit internal structure from the inserted language. The presence of other-language surface level structure in inserted multimorphemic units may indicate activation of the other language for the duration of the switch. The presence of a different grammar from that of the rest of the sentence, however, should not automatically be seen as evidence that the insertion itself has been derived compositionally (i.e. assembled morpheme by morpheme). In his analysis of a Turkish-Dutch corpus Backus (2003) argues that most other-language multimorphemic elements in CS are inserted into speech as “chunks” or lexical units. He argues that frequent combinations of morphemes may become entrenched in the speaker’s mental lexicon, and be stored and produced as units, rather than being assembled compositionally. In many cases, the units may possess an “idiomatic meaning”, i.e. a semantic interpretation which can be slightly or radically different from the mere sum of the elements required to assemble the unit.

If multimorphemic elements are indeed stored as units in the lexicon, then at least some complex NPs would be inserted in speech not dissimilarly from single nouns, meaning that there would be no reason to treat the two categories as distinct. In the case of examples (49) and (50), it may be difficult to argue that much idiomaticity is present, since the semantic interpretation of the NPs can easily be achieved compositionally. There are other NPs in the data, however, which could more easily be interpreted as unitary insertions. One such example is given in (51).

51. \*MAR: si sono incontrate in una summer@en school@en.  
REFL be-PRS.3PL meet-PST.PTCP.F.PL in DET

*'they met each other at a summer school.'*

[Manchester 3]

The inserted NP *summer school* has at least some degree of non-compositional meaning. While it is true that the interpretation of 'a school that teaches during the summer' can be derived by assembling the meaning of the two components *summer* and *school*, the compositional interpretation does not convey the idea that summer schools normally teach students at university level, have very specialised subjects, often take the form of workshops etc. All of these elements are not discernible from a compositional interpretation of the NP, but are known by speakers as part of a more idiomatic (i.e. non-compositional) interpretation of the lexical item. *Summer school* could therefore be regarded as an example of inserted multimorphemic unit.

## 6.2.2 Adjective insertions

Adjectives in English can appear either as attributive or predicative. Predicative adjectives are typically used in conjunction with a verb, most frequently a copula (as *red* in the sentence *the shirt is red*). Attributive adjectives, on the other hand are produced as part of a nominal construction, and normally immediately precede the noun they refer to (as *black* in *the black shirt is in the wardrobe*).

Placement of attributive adjectives in English is rigidly constrained. Also, both attributive and predicative adjectives in English are invariable, not showing agreement in gender or number with their referent.

As in English, adjectives in Italian can occur both as predicates and attributes. Predicative adjectives behave very similarly to those in English, and are produced in conjunction with a verb, most frequently the copula *essere* 'to be' (e.g. *rossa* 'red' in *la camicia è rossa* 'the shirt is red'). Attributive adjectives, on the other hand, are placed adjacent to the noun they refer to (as *bianca* 'white' in *la camicia bianca è nell'armadio* 'the white shirt is in the wardrobe'). The two uses of adjectives exist in both languages and their different functions are encoded through similar means, i.e. adjacency to referent vs. co-occurrence with a copula. Differences between the two languages arise with reference to agreement. While English adjectives are invariable, Italian adjectives, both attributive and predicative are inflected in gender and number to agree with their referent. Agreement is shown in the word's inflection, which encodes both gender and number. Compare *la camicia è rossa* 'the shirt is red', where the adjective's ending *-a* expresses agreement in gender and number with singular feminine *camicia*, with *i pantaloni sono rossi* 'the trousers are red', where the adjective ends in *-i* to mark agreement with masculine plural *pantaloni* 'trousers'. A further difference between English and Italian is in the position of attributive adjectives. While English shows a strict modifier-head order, the position of attributive adjectives in Italian is less constrained, and both pre and post modification of a noun are possible.<sup>30</sup> The two languages, therefore, are maximally divergent with regard to attributives, where both morphological and syntagmatic differences occur. For predicative adjectives, on the other hand, differences are limited to the presence or absence of agreement.

---

<sup>30</sup> Comparative corpus data show that, at least in the written medium, Italian allows greater freedom for adjectival placement than other Romance languages (Radatz and Stammerjohann, 1996) It would be incorrect, however, to assume that all attributes can occur either pre or post-nominally. Some categories of adjectives have relatively fixed positions when used as attributes, mostly based on whether the adjective expresses an objective quality or a subjective evaluation of the noun that it refers to. Some attributive adjectives also accompany the change of position with a shift in meaning.

In the present data, adjectives sometimes appear with an attributive function as part of multiword NP insertions, as in examples (49) and (50) in section 6.2.1.1 above. Overall, insertions of single adjectives – attributive or predicative – are quite rare in both data sets. As mentioned earlier, this is particularly rare in the Milan data. Only 6 Italian adjectives in an otherwise English utterance were found in the Milan data, against 23 English insertions into Italian in Manchester. In absolute terms, insertion of single adjectives is three times more frequent in Manchester than in Milan.

In both data sets inserted predicative adjectives are much more likely to occur than attributive ones. In fact, there is only one occurrence of an unambiguously attributive adjective in the Manchester data (reproduced in example (62), and none in the Milan data.

The greater ease of insertion for predicative adjectives compared with attributive ones could be motivated in terms of congruence (see Deuchar, 2005). Analysing Welsh-English CS data, Deuchar argues that the low number of inserted adjectives in the data set may at least partly be due to lack of congruence between the languages with regards to placement of attributive adjectives. The same situation holds in our data.

Below are two examples of insertions of predicatives from Manchester (52) (presented earlier as example (20)) and Milan (53):

52. \*NIC: il mio esempio è biased@en &=laughs .  
DET my.M.SG example be-PRS.3SG

*'my example is biased'*

[Manchester 4]

53. \*KEL: oh@0: # she's so paurosa@it .  
easily.scared

[Milano 2]

In (52) *biased* appears with no overt marking, consistently with what would be expected if the adjective had been used in an English monolingual context. Interestingly enough, however, in this case this is not in overt conflict with the rules of Italian. While it is true, as mentioned earlier, that adjectives agree in

gender and number with their referent, adjectives not ending in an unstressed vowel are normally invariable. This group includes both adjectives ending in a stressed vowel, such as *blu* 'blue', or ending in a consonant. Examples of the latter tend to be established borrowings such as French *chic* or English *snob*, neither of which would be marked for agreement if used in a monolingual Italian context. While a consonant cluster such as the one in *biased* is a particularly unusual ending for Italian phonotactics, the lack of overt agreement on the insertion is perfectly acceptable for the rules of Italian. If we were to measure the integration of the adjective into Italian based on the presence of overt Italian morphology, we would not have enough elements here, due to the case of isomorphism between the two languages and the invariability of the adjective. In (53) on the other hand, *paurosa* is inflected in the singular feminine, complying with Italian rules, so as to agree with its referent *she*. Since English does not mark adjectival gender-number agreement, this could have only come from Italian. However, since adjectives in Italian do not have a zero form such as *\*pauros* (unless one assumes that the masculine singular ending is the default), the speaker is forced to produce it with some ending. In this case, she may have selected feminine singular ending *-a* based on semantic criteria alone.

### 6.2.3 Verb insertions

Verb insertions are very rare in both data sets, involving only a handful of tokens. However, the scarcity of inserted verbs should not be surprising in the particular context at hand, given the morphological and syntactic differences between the two languages. Indeed, the rarity of the examples could almost lead one to consider the occurrence of verb insertions as a feature of the idiolect of particular speakers (only seven out twenty speakers inserted verbs in the two groups). While the limited number of tokens prevents us from making strong generalisations, some observations can be made about trends emerging from the instances encountered in the data. One is that verb insertions seem to occur at points where a linear equivalence between the two languages can be observed, as in example (54), which is taken from Manchester:

54. \*SAR: no sto &do [//] sto uhm@0:  
no stay-PRS.1SG stay-PRS.1SG IM  
struggling@en con le conclusioni .  
with DET conclusions

*'no I'm I'm uhm struggling with the conclusions.'*

[Manchester 1]

In (54) the speaker inserts the verb *struggling* after encountering problems in the formulation of the utterance. The insertion is clearly flagged, as it is preceded by retracing and the hesitation marker *uhm*, both indicating that a problem at some level in the production of the utterance has occurred. What we are interested here in, however, is the linear equivalence between the English and Italian monolingual versions of the sentence, as illustrated in (55) below:

55. I am struggling with the conclusions.  
(lo) sto facendo con le conclusioni.

Both Italian and English express the present progressive through a periphrastic construction (although Italian can sometimes substitute the periphrasis with a present simple). In both languages the periphrasis is composed of an aspectual auxiliary followed by a non-finite form of the main verb – what is traditionally called a gerund. The mixed clause produced by the speaker seems to be the result of a compromise strategy, with the progressive aspect expressed by both the Italian auxiliary *sto* and the English *struggling*. Overt conflict is thus avoided by switching at a point where the structures of the two languages are isomorphic. A second observation concerns the preference for non-finite forms over finite ones. This is true of example (54) above and (56)<sup>31</sup>, (58) and (59) below. In fact, no insertions of finite forms of verbs are found in either data set.

---

31 For the sake of clarity, the example, which appears in a rather long and complex sentence, is reproduced only as a fragment. The full sentence is reproduced below:

\*LUC: quindi ehm: do un' interpretazione uh in chiave uh allegorica eccetera <ai ai ai ai> [//] ai dipinti eccetera che vengono # uh: **referenced** # <uhm no da un punto di vista appunto:> [//] <uhm come dire da un> [//] dandoci una chiave di lettura del: [//] gotica # eccetera.

*'so uhm I'm giving uh an allegoric interpretation to the paintings and so on which are uh referenced uhm you know from a point of view uhm, how to say, by interpreting them from a gothic point of view and so on.'*

<b>56.</b> *LUC:	dipinti	eccetera	che	vengono #
	paintings	etcetera	REL.PRN	come-PRS.3PL
uh:@0	referenced@en # <	uhm@0	no .	
IM		IM	no	

*'paintings etcetera that are uh referenced uhm no'*

[Manchester 2]

While in (54) a present participle is inserted, in (56) a past participle is found as part of a passive construction. The verb *venire* 'to come', is used as an auxiliary in Italian to construct the passive voice for simple tenses, as an alternative to *essere* 'to be'.

In (56) again, the insertion occurs at a point of linear equivalence, as shown below:

<b>57.</b> paintings	that	are	referenced.
dipinti	che	vengono	citati.

The past participle *referenced* is inserted at an equivalence point between the two languages, but it follows the grammar of its source language in that it does not agree in gender and number with its subject, as would be required of the Italian construction (the Italian equivalent *citati* ends with the masculine plural ending *-i*, thus agreeing with the subject *dipinti*). As a matter of fact, the Italian past participle in passive constructions behaves exactly like an attributive adjective, requiring the same agreement. As discussed in the previous section, both English and Italian past participles can be used as verbal adjectives. Unlike adjectives, however, there are strict phonological constraints governing Italian past participles used as verbs. All past participles must obey a precise template. The exact rules are rather complex, but for our purposes it will suffice to say that all past participles must end in either *-t(o)* or *-s(o)*, with the final vowel changing to mark gender/number agreement. The absence of overt Italian marking on *referenced* and the presence of English past participial suffix *-ed*, therefore, shows lack of morphological integration of this insertion into Italian.

A different kind of insertion is shown in examples (58) and (59) below:

58. \*VIT: visto che lo puoi:  
 see-PST.PTCP.M.SG COMP it-OBJ.PRN can-PRS.2SG  
 scollare@s:en+it molto velocemente .  
 scroll-INF very quickly

*'given that you can scroll it very quickly'*

[Manchester 5]

59. \*GIU: però fatto sta che un po' ci  
 but fact stay-PRS.3SG COMP DET bit REFL.PRN.1PL  
 siamo: smailicchiati@s:en+it ma uhm:@0 # niente:  
 be-PRS.1PL PRT.MAIL-PST.PTCP.M.PL but IM nothing

*'but the thing is that we've emailed each other for a while but uhm nothing'*

[Manchester 3]

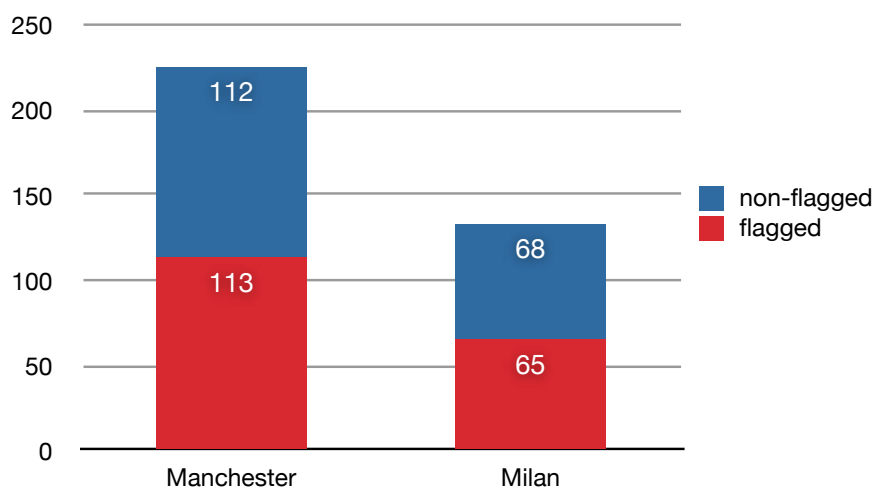
In both cases, non-finite mixed forms are inserted into an otherwise Italian clause. Both are flagged by the lengthening of the preceding word. Unlike (54) and (56), however, the English verbs *scroll* and *mail* have been integrated by means of derivational morphology. In (58) the infinitival verbal suffix *-are* is added to the English verb – an infinitive is what would be expected in Italian after the modal verb *potere*. The use of the infinitival ending *-are* to form verbs is a productive derivational process in Italian, and is frequently used with both native Italian neologism and loanwords. A recent example which is also attested in Italian dictionaries is *chattare* 'to talk in an online chat', from English *chat*. In (59), a more complex form of integration occurs. Inflectional morphology is added to make the verb *mail* into a past participle, and is also made to agree in gender and number with the subject of the verb. Some evaluative morphology (see Gambino, 2010), however, is also added (initial *s-* and *-icchi* before the aspectual *-at* of the past participle). These extra affixes modify the verb by adding both a diminutive/attenuative meaning and a frequentative aspect to the verb, following a morphological process that is productive in contemporary Italian and affects verbs as well other grammatical categories. Insertions such as (58) and (59) are different from (54) and (56) not in terms of flagging (which is present in all four instances) but with regards to the presence or absence of morphology from the host language on the insertion.



### 6.3 Analysis of flagging

In this section, the occurrence and types of flagging with different insertions are analysed in detail. The relation between flagging and frequency is explored both for insertions in the aggregate and for single grammatical categories. Lone nouns are analysed in greater detail than other categories, due to the number of tokens being considerably higher. A direct comparison is made between nouns in bilingual and monolingual contexts with reference to the occurrence and type of flagging and related discourse phenomena (in 6.3.1.1).

A first count of flagged vs. non-flagged other-language items in the two data sets gives the results shown in **Figure 3** below:



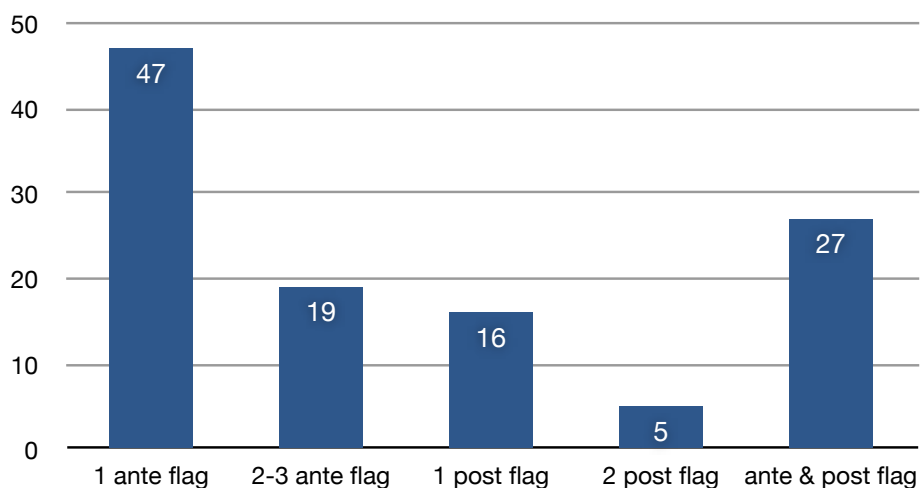
**Figure 3.** Count of flagged vs. non-flagged switches in the two data sets.

While the absolute difference in the number of switches between the groups is rather obvious, similarities emerge between the two groups in the percentage of flagged switches, which amounts to 50% of inserted material for Manchester and 49% for Milan (red section of the bars).

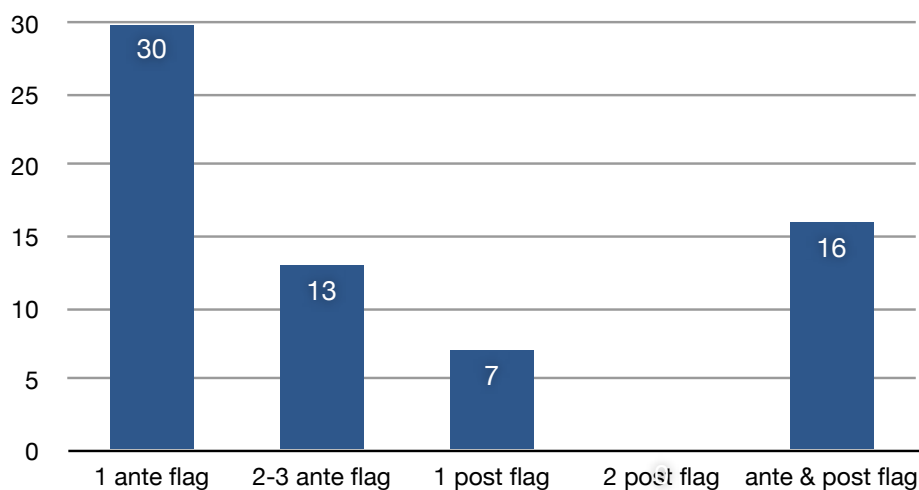
After the first count, all flagged items were classified according to the position and the number of flagging devices produced in their immediate proximity.

Following Hlavac (2006) an explicit distinction was drawn between flags preceding a switch and flags produced after the switch. The two categories are

labelled here 'ante' and 'post' flag respectively. Bearing this distinction in mind, the two basic categories are further subdivided according to the number of flagging devices used. One third category not present in Hlavac is also added, i.e. that of inserted items showing flagging both to their left and their right ('ante & post flag'). The results from this classification of flagged items are shown in **Figure 4** and **Figure 5**.



**Figure 4.** Position and number of flags in Manchester data.



**Figure 5.** Position and number of flags in Milan data.

The two groups show very similar patterns with regards to the positions and quantity of flagging items, except with the ‘two post-switch flags’ category, which only appears rather marginally in the Manchester data. No statistically significant difference was found between the position and number of flagging devices between the two groups.

Two general observations can be made: one is that one single flag before the switch seems to be the most common flagging strategy adopted by speakers in both groups; the second is that there is a general preference for flags to be produced before the switch rather than after. A partial exception is represented by the flags classed as explicit justification and translation equivalents, which have a tendency to occur after the switch. The example in (60) shows an instance of a post-switch flag through a translation equivalent.

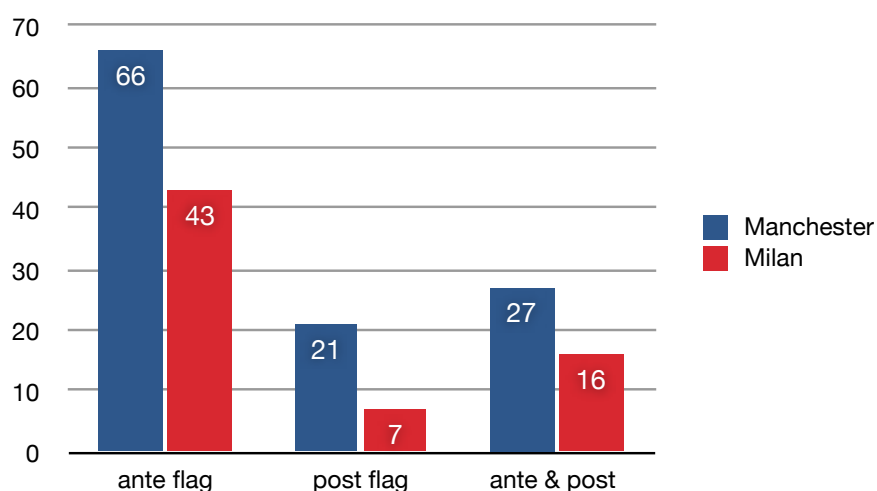
<b>60.</b>	*PAO:	e	poi	la	questione	delle: [/]	delle:	
		and	then	DET	issue	OF.DET	OF.DET	
	bad@en	words@en	delle: [/] #	delle:	parolacce #	e:	i	colloquialismi
			of.DET	of.DET	bad.words	and	DET	colloquialisms

*‘and then the issue of the [/] of the bad words of the [/] # of the bad words # and colloquialisms.’*

[Manchester 2]

In the example it is possible to see the occurrence of *parolacce* ‘bad words’, a literal translation equivalent of *bad words*, after the insertion itself. To be precise, the inserted NP *bad words* is flagged to its left as well as to its right by retracing and lengthening of vowels. It could also be argued that the insertion itself is part of a larger retracing sequence *parolacce* ‘bad words’, a literal translation equivalent, after the insertion itself.

The higher frequency of flagging before the switch can be seen more clearly if the categories used in **Figure 4** and **Figure 5**: are collapsed to only three more general ones, with no distinction according to the number of flags. The resulting categories would simply be pre-switch flags only, post switch flag only and both pre and post switch. The results are illustrated in **Figure 6**. While the picture that emerges is clearer, however, no statistical significant difference is found between the groups.



**Figure 6.** Flagged insertions by position of flags.

### 6.3.1 Flagging of single nouns

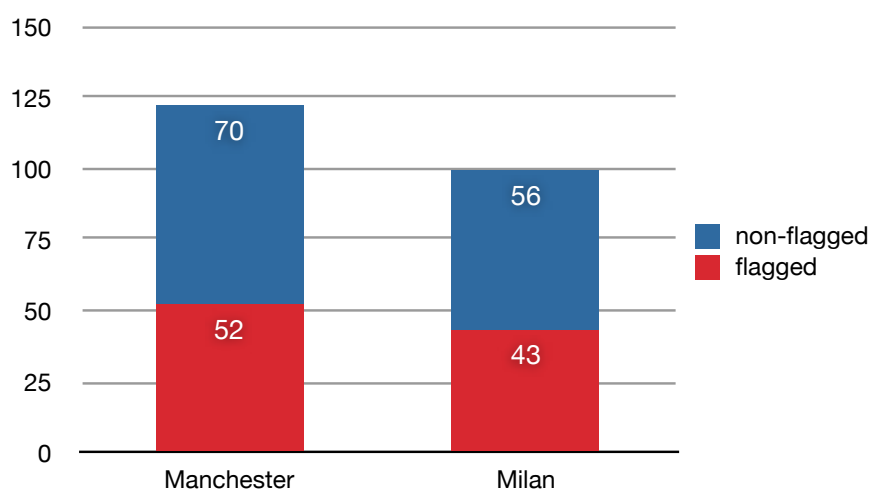
Since single nouns are the category with the highest number of insertions in the data, it is possible to carry out some more fine-grained analysis of flagging for this morphological category. As illustrated in **Figure 2** in section 6.2, there are 122 inserted noun tokens in Manchester and 98 in Milan. The count of noun types gives 59 for Manchester and 63 for Milan (see **Table 2**). This translates into a type/token ratio of 0.48 for Manchester and 0.64 for Milan. This can be interpreted as showing that there are more types in Milan with only one token, whereas in Manchester, more types are found occurring more than once.

	Manchester	Milan
noun tokens	122	98
noun types	59	63

**Table 2.** Noun tokens and types.

As illustrated in **Figure 7**, quite strikingly, the two data sets present the identical rates of 43% flagged nouns (red) and 57% non-flagged nouns (blue). As a morphological category, nouns tend to be inserted in speech with slightly less flagging than other types of insertions – the general flagging rates are 50% for Manchester and 49% for Milan. It should be noted that the lexical items being

the most easily borrowable part of the language in contact situations *viz à viz* morphological and syntactic features (see Sankoff, 2001), and nouns typically being the most borrowed part of the lexicon, it is perhaps unsurprising that other-language nouns are inserted in speech production somewhat more smoothly than elements belonging to other morphological categories. Still, the flagging rates are quite high compared with similar studies. Turpin (1998) finds only 9% of English-origin nouns flagged in English discourse, while Jones (2005) finds that 23% of inserted English items in Jersey Norman French are flagged across all grammatical categories. Jones interprets the relatively low rate of flagging as a sign of the degree of penetration of English contact forms in Jersey French, due to the length of contact between the two languages. This difference between her data and the present set could also be in part due to the fact that different phenomena (such as lengthening of syllables) have been included in the category of ‘flagging’ in the present study.



**Figure 7.** Rate of flagging for nouns in the two data sets.

### 6.3.1.1 Establishing a baseline for flagging of nouns

The majority of the phenomena gathered here under the umbrella term ‘flagging’ are not exclusive to bilingual data. Hesitations, pauses, reformulations and other discourse phenomena are also routinely produced in monolingual speech, as

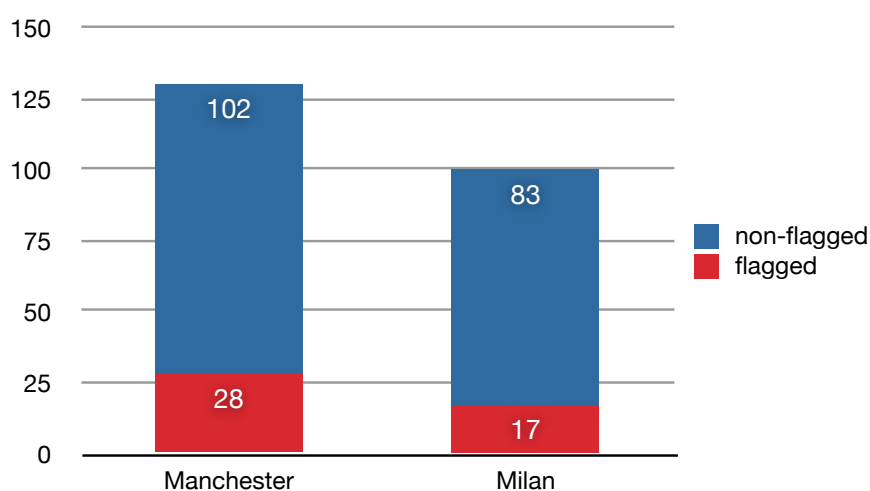
discussed in section 3.3. It is therefore important to establish whether the flagging co-occurring with other-language items is quantitatively and/or qualitatively different from the flagging normally co-occurring in monolingual speech. In order to obtain a fair comparison, the speech of the same speakers should be used when measuring rates for both code-switched and monolingual speech. In their study on the conversational functions of CS versus monolingual speech Gardner-Chloros *et al.* (2000) note how the comparison of naturalistic bilingual and monolingual speech from the same speakers poses considerable methodological problems. In order to have a maximally fair comparison, the monolingual and bilingual production of the same speakers would have to be collected on separate occasions. The researcher would have to give participants explicit instructions about how to speak in each recording, therefore having a negative impact on the naturalness of the collected data. Alternatively, bilingual speakers would have to be paired with monolinguals to obtain a sample of monolingual speech, which would result extremely time-consuming. To avoid these problems, Gardner-Chloros *et al.* suggest “tak[ing] advantage of the fact that bilinguals are also, by definition, monolinguals at the same time, and in the same conversation may at times resort to codeswitching and at other times speak monolingually” (2000: 1312). The monolingual stretches in bilinguals’ speech, they argue, can be used to make a comparison with the more code-switched sections. In this way, the same participants in the same context are used, effectively minimising the role of external factors that could affect their speech. In the present data, the comparison between monolingual and bilingual speech was quite easy. Instances of language alternation are rather scarce and scattered in the corpus, resulting in many long stretches of what could reasonably be regarded as monolingual speech.

#### 6.3.1.2 Quantitative differences

In order to check whether there was a difference between the flagging of single monolingual nouns and switched ones, a random sample of monolingual nouns was extracted from the data. The section of the corpus that had been transcribed

in its entirety was used to extract the data<sup>32</sup> (see section 5.2). In each transcript, each tenth text line was searched for a single noun, which was then transferred onto a spreadsheet together with any information regarding its flagging. If the line contained no nouns or if the noun was switched, the tenth next line would be used. Proper nouns and nouns that were part of multiword NPs that were deemed to have idiomatic meaning as a unit (e.g. English *point of view*) were also excluded from the analysis. Nouns that were marked as ambiguous (i.e. in the @0 category), on the other hand, were included in the sample.

Using the above method, 130 Italian noun tokens were extracted from the Manchester data to match the 122 English-origin nouns. In the Milan data, 100 nouns were extracted to match the 98 Italian insertions. The rates of flagging for the two samples are represented in **Figure 8**.



**Figure 8.** Flagging of monolingual single nouns.

The rate of flagging is considerably lower for monolingual nouns in both data sets, with 17% flagged tokens in Milan and 22% in Manchester, against 43% for other-language nouns in both data sets.

A chi-square test was conducted on the switched and monolingual nouns in both data sets. The difference in the percentage of flagging is significant both in the

---

<sup>32</sup> The conversations used are **Manchester 1**, **Manchester 2**, **Manchester 4**, **Milano 2**, **Milano 3** and **Milano 5**. **Manchester 4** is only partially transcribed. The nouns used for the sample only come from the section that has been fully transcribed.

Manchester data ( $\chi^2=11.96$ ,  $df=1$ ,  $p= .0005$ ) and in the Milan data ( $\chi^2= 15.28$ ,  $df = 1$ ,  $p<.0001$ ). What these results mean is that if an other-language single noun is inserted in speech, it is significantly more likely to be flagged than a single noun in the same language of the surrounding speech. In other words, with regards to the occurrence of flagging, other-language nouns appear to pattern differently compared with monolingual nouns in the same data.

### 6.3.1.3 Differences in types of flagging

Having ascertained that the presence of flagging is significantly higher for other-language (switched) nouns in each data set than for same language nouns, the next question to ask would be whether there are any significant differences between flagging of inserted nouns and that of host language nouns in terms of flagging strategies. That is to say, whether the types of flagging used for other-language nouns are different from those that appear with same-language nouns. In order to answer this question, the switched flagged nouns would have to be compared with the non-switched flagged nouns, and the type of flagging used for each recorded. Since non-switched flagged nouns (i.e. English flagged nouns in Milan and Italian flagged nouns in Manchester) are quite rare in the data, a larger sample had to be used for this analysis, so as to have roughly the same number of tokens in the switched and in the non-switched category. The sample from the previous analysis was therefore expanded by extracting lone nouns in every fifth line from the conversations that had been fully transcribed<sup>33</sup>. The new sample consisted of a total of 234 Italian nouns in Manchester and 204 English nouns in Milan. In the Manchester set, 50 nouns presented some flagging, while in Milan there were 41 flagged nouns. The flagging strategies used for these non-switched nouns were then compared with those that appeared with the switched nouns in the same data set. For nouns showing more than one type of flag, only the first occurrence of flagging was counted, since a single flag was the prevalent strategy used for nouns. Also, no distinction was made between pre and post switch flags,

---

<sup>33</sup> See footnote 32 for the specific conversations used for the sample.



since there was no statistical significant difference between the groups in terms of position of flags, as illustrated in section 6.3 (see also **Figure 6** in the same section). The results from the new analysis are shown in **Table 3**.

	MANCHESTER		MILAN	
	non-switched	switched	non-switched	switched
lengthening	11	17	4	6
LFP	0	0	2	4
metalinguistic	1	5	0	8
NLFP	4	5	2	7
paralinguistic	3	1	1	1
retracing	13	13	7	5
UP	18	11	25	13
<b>TOTAL</b>	<b>50</b>	<b>52</b>	<b>41</b>	<b>44</b>

**Table 3.** Types of flagging in non-switched and switched nouns.

The table shows how flagging for nouns in both corpora is carried out using a wide variety of devices. Because of the size of the sample and the number of flagging strategies resulting in many single figure cells, it was only possible to test for any variation between the two groups in the overall distribution of different types of flagging, rather than for significant variation between single types of flagging.

A Fisher's test was thus carried out comparing overall flagging patterns for non-switched vs. switched nouns in Manchester and Milan. The results show that there is a statistically significant difference in the types of flagging used in Milan ( $P = 0.006$ ), meaning that the flagging devices used for Italian switched nouns show a different distribution overall from non-switched English nouns. Unfilled pauses (UP), in particular, are much more likely to appear with English nouns (25 tokens) than they are with Italian nouns (13 tokens). Conversely, metalinguistic commentary (which includes translation equivalents, explicit pre-empting and/or justification of the insertion) is much more likely to occur with an Italian noun insertion (8 tokens) than with an English non-switched noun (no tokens). The higher presence of metalinguistic commentary in conjunction with inserted Italian nouns could be seen as indicative of the fact that speakers are particularly conscious of inserting an other-language item in their speech. Other types of

flagging, such as lengthening, retracing and filled pauses, on the other hand, only show modest variation between switched and non-switched contexts. This indicates that the use of these strategies is not to be linked exclusively to flagging of switched items.

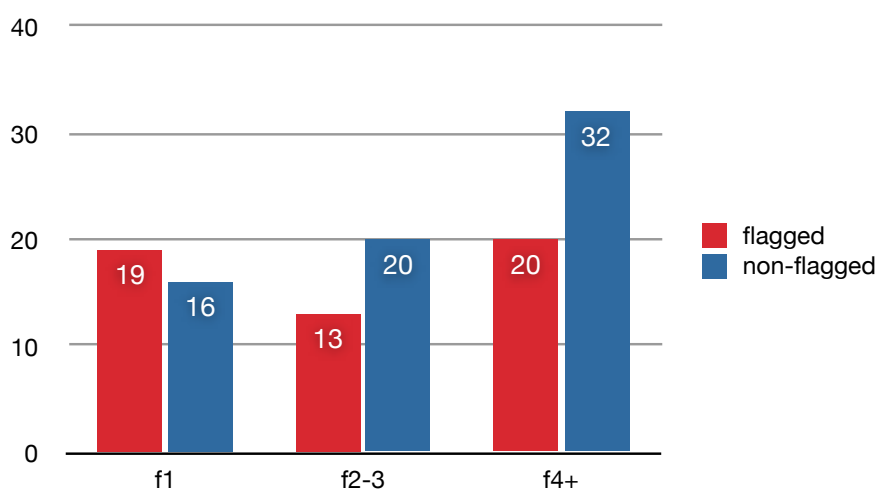
In Manchester, on the other hand, the difference in the distribution of flagging types between the groups is not statistically significant. Some categories, such as lengthening and unfilled pauses, show a visible increase or decrease for inserted English nouns vs. non-switched Italian nouns, but no dramatic variation can be observed.

It is worth noting that, while not much can be said from a statistical point of view, some differences do emerge between overall types of flagging in Manchester vs. Milan, regardless of the type of item. There seems to be a preference for lengthening, retracing and UPs in Manchester, while in Milan, only the UPs are relatively frequent. Further, the category of LFPs (words such as *like* in English) only appears in the Milan data. Markers fulfilling a similar function also exist in Italian (e.g. *tipo*, 'like' or *cioè* 'that is'), but they do not appear to be as frequent as their English counterpart, at least not in the present data. This could also explain the different rates of LFP in the two groups.

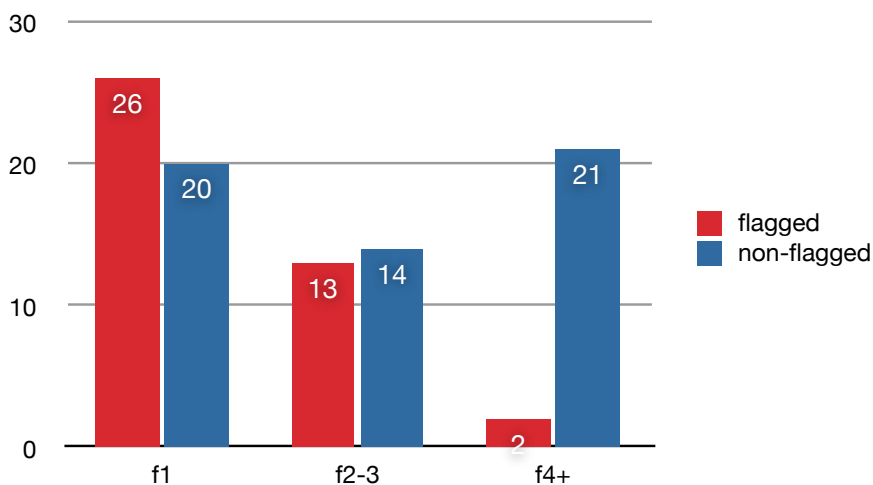
General differences between Manchester and Milan may be ascribed to general differences between English and Italian. Monolingual speakers of the two languages may use different devices and discourse markers to signal problems in speech, and this could be reflected in the speech of bilinguals. Lengthening of syllables, for instance is much more commonly found in the Manchester data, possibly as an extension of a hesitation strategy existing in Italian. Statistical analysis, however, shows that there is no significant difference between the groups in the distribution of flagging types for non-switched nouns. A not very strong, but nonetheless significant difference is found instead when the distributions of flagging types for switched nouns in the two groups are compared ( $P = 0.0258$ ).

### 6.3.1.4 Flagging and frequency: first analysis

As mentioned earlier, aside from the absolute numerical difference, the two data sets show different type/token ratios for nouns. In Milan, in particular, more types are present which occur only once across the corpus (35 in Manchester against 46 in Milan). Since the flagging rates are identical at a first analysis, the data were then reanalysed by dividing the nouns according to their frequency in the corpus. They were subdivided arbitrarily in three categories: noun types with just one token (f1), with two to three tokens (f2-3) and with four or more tokens (f4+). The flagging rates were counted again, giving the results in **Figure 9** and **Figure 10**.



**Figure 9.** Noun flagging and frequency in Manchester.



**Figure 10.** Noun flagging and frequency in Milan.

In the Manchester data (**Figure 9**) it is possible to observe an overall decrease of flagged nouns compared with unflagged ones as frequency increases. While the split is almost even in the f1 category, flagging seems to diminish in the f2-3 and f4+ groups. In absolute terms, however, even high-frequency items show a considerable rate of flagging, with 20 flagged tokens against 32 non-flagged ones. This means that even some widespread nouns used by several speakers will be produced at least once with flagging.

The situation in the Milan data (**Figure 10**) is remarkably different, with a dramatic effect of frequency of items on their flagging rates. While in the f1 column the flagged/unflagged results are similar to the Manchester group, and in the f2-3 column the rates are almost identical, a very abrupt decrease is seen in the f4+ category, where flagging all but disappears. Even in absolute terms a clear trend can be observed, with the number of flagged tokens decreasing from 26 in f1 to 13 in f2-3, to 2 in f4+. This is in clear contrast with the situation in Manchester, where in terms of absolute quantities, flagging remains somewhat constant, with 19 tokens in f1, 13 in f2 and 20 in f4+.

The two groups of speakers are therefore showing rather different behaviour with regards to the flagging of nouns. If the presence of flagging is seen as a way to draw attention to the switch or to show that speakers are conscious of inserting an item from a different language, the results for nouns show that, at least in Milan, with the increase of an item's frequency speakers tend not to draw attention to it and switch smoothly in conjunction with the insertion. A Fisher's test between the flagging rates in the f4+ category shows that the difference between Manchester and Milan is statistically significant ( $P = .01$ ). This is not the case for inter-group comparisons in the f1 and f2-3 categories, where the observable difference is not statistically significant.

The different behaviour of items in the f4+ could be seen as evidence that, at the discourse level, high-frequency Italian insertions into English pattern more like English nouns in the speech of Milan speakers, thus becoming more similar to non-switches (see **Figure 8** in section **6.3.1.2**). A significant difference is also found between f1 and f4+ items in Milan ( $P = .0002$ ), reinforcing the idea that the occurrence of a flagged insertion is connected to the item's frequency.

In Manchester, on the other hand, a different picture emerges: while it is possible to say that the frequency of an item has some effect on the appearance of flagging, this effect seems to be much weaker than in Milan, and is not statistically significant. This is apparent in the f4+ category, in which more than a third of tokens is produced with flagging of some sort, compared with Milan, in which only two out of twenty-one nouns are flagged.

### 6.3.1.5 Flagging and frequency: reclassifying the data

The analysis in section 6.3.1.4 shows a significant difference between groups restricted to the f4+ category and within the Milan group between the f4+ nouns and those with only one occurrence. Since the size of the sample is quite small, however, it is possible that even items with an absolute relatively high frequency in speech appear only once in the data. The frequency bands used in 6.3.1.4 were selected to provide a more accurate picture of the distribution of the data. However, they may also be too sensitive for the size of the sample. It would therefore be possible to draw separations along different lines in a way that better suits the type of data under analysis.

The noun insertions were therefore reclassified according to frequency using only two categories: f1 for items with only one occurrence and f2+ for items occurring more than once. This was done to better suit the size of the sample, in which even relatively frequent words may present only few tokens. The results of this recount are shown in Figure 11 and Figure 12.

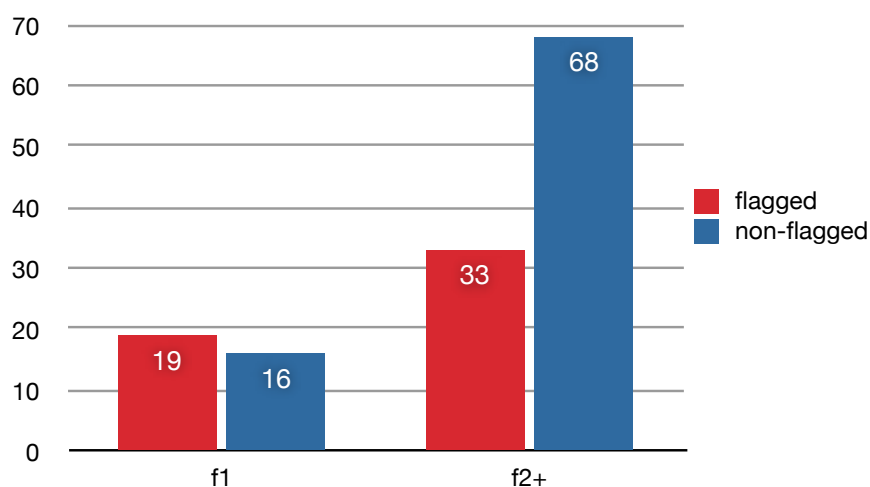
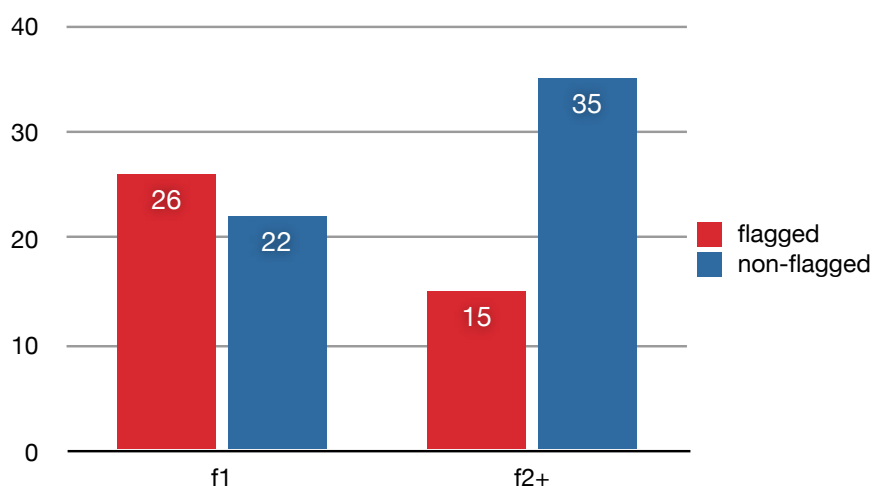


Figure 11. Noun flagging and frequency in Manchester (recount).



**Figure 12.** Noun flagging and frequency in Milan (recount).

If the data are reclassified into only two categories, the patterns in either group become remarkably similar. Any statistically significant differences between the groups are lost, and the dramatic drop in the rate of flagging in Milan for the f4+ category is no longer apparent. A significant difference between f1 and f2+ nouns, however, emerges within each of the data sets. In Manchester the difference is only marginally significant ( $\chi^2 = 4.2666$ ,  $df = 1$ ,  $p = .0389$ ) while it is more robust in Milan ( $\chi^2 = 4.9266$ ,  $df = 1$ ,  $p = 0.0265$ ). These results reinforce the idea that a relationship holds between the frequency of the item in the data and its production with flagging. As already shown by the previous analysis this effect is stronger for speakers in Milan.

### 6.3.2 Flagging of multiword noun phrases

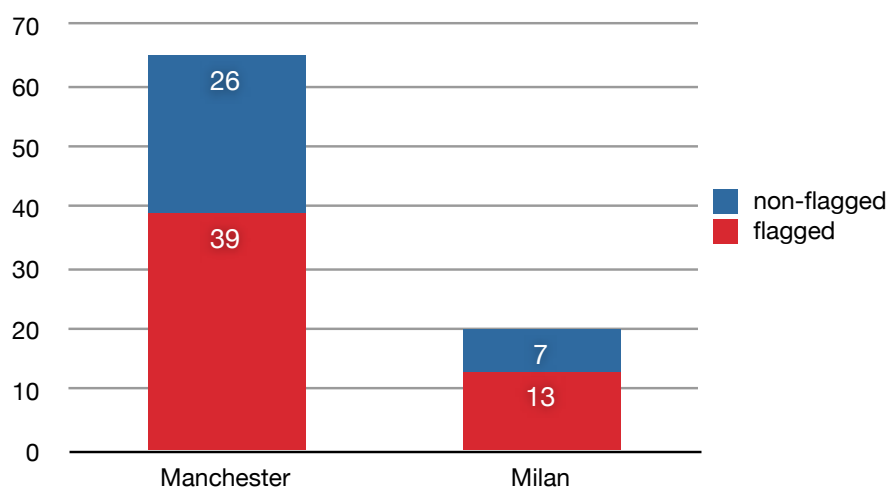
The numbers for inserted multiword NPs are rather lower than for lone nouns; however, it is still possible to make some general observations about this type of insertion.

In the Manchester data 65 instances of multiword NP insertions were found, compared with only 20 in Milan. The difference between the two data sets with regards to the number of NP insertions is significant ( $\chi^2 = 8.1091$ ,  $df = 1$ ,  $p = 0.004$ ). This could be interpreted as meaning that the production of English

multiword NPs in Italian discourse (Manchester) is somewhat facilitated compared to Italian multiword NPs in English discourse (Milan). It is equally plausible, however, that this is simply due to differences in topic and or register between the groups, which could have resulted in different types of inserted material.

The flagging rates for both groups are illustrated in **Figure 13**. As was the case for lone nouns, the percentages of flagged NPs are quite similar although not identical (60% in Manchester and 65% in Milan).

The main difference emerges when comparing the rates in each group with those obtained for lone nouns, where only 43% of tokens in either group appeared with flagging. This observable difference appears to be statistically significant in Manchester ( $\chi^2 = 4.4537$ ,  $df = 1$ ,  $p = 0.0348$ ), but not in Milan. At least in the Manchester data, therefore, it could be argued that for multiword nominal constructions there is a significant increased rate of flagging. This does not seem to be the case for speakers in Milan, who do not produce significantly more flagging when inserting complex Italian NPs in English than when inserting single Italian nouns.



**Figure 13.** Rate of flagging for multiword NPs in the two data sets.

### 6.3.2.1 Flagging and frequency

In sections 6.3.1.4 and 6.3.1.5 it was shown how a relationship holds between the overall frequency of an inserted lone noun in the corpus and the appearance of flagging. In both groups, switch types that only had one occurrence were significantly more flagged than nouns with multiple occurrences. When the data were divided in several fine-grained categories for different frequencies the overall effect was lost. This difference became visible when the data were divided only in the f1 (one occurrence) and f2+ (more than one occurrence) categories. The same classification used for nouns was tested on NPs, in order to test whether more frequent items were produced more smoothly than less frequent ones. Inserted NPs from the two groups data were subdivided in the f1 and f2+ categories and the respective flagging rates compared. On the whole, however, the majority of inserted NPs occurred only once in the data.

In Manchester (**Figure 14**) it is possible to see a dramatic difference between the f1 and f2+ category. While in the f1 category the overwhelming majority (83%) of insertions is flagged, in f2+ the rate decreases to just half of the total tokens. The difference between the groups is highly significant ( $P < .0001$ ). This means that English NPs with a higher frequency in the data are inserted with much less hesitation into Italian discourse than those appearing only once.

For the Milan data (**Figure 15**) a disclaimer is necessary. The number of inserted NPs is very low and all insertions occur only once, with one exception. In **Milano 3** one type is found (*scuola materna* 'kindergarten') which appears four times. A statistical test was therefore not possible on the Milan data, since the f2+ consisted of different tokens of only one NP. It is perhaps worth noting that, if we focus on the f1 category, inserted Italian NPs in Milan speech show a somewhat similar pattern of flagging to f1 in Manchester. 11 flagged NPs are found against only five non-flagged ones, as shown in **Figure 15**. Not much can be said for the f2+ category, except noting that the NP *scuola materna* is produced twice with flagging and twice with no flagging. One of the flagged occurrences is reproduced in example (61), where the insertion is preceded by both *like* (a LFP) and lengthening on the determiner *the*.



61. \*AMI: is it the whole complex with his like the: **scuola@it** **materna@it** [>] .  
kindergarten

[Milano 3]

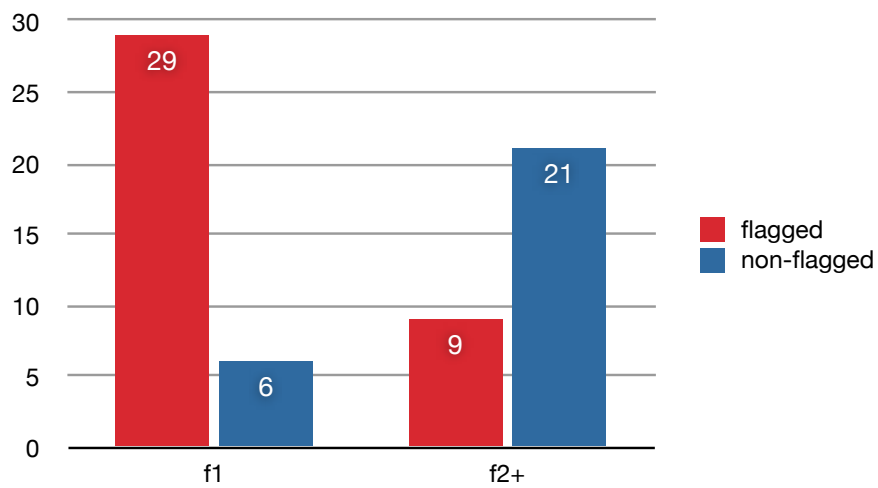


Figure 14. NP flagging and frequency in Manchester.

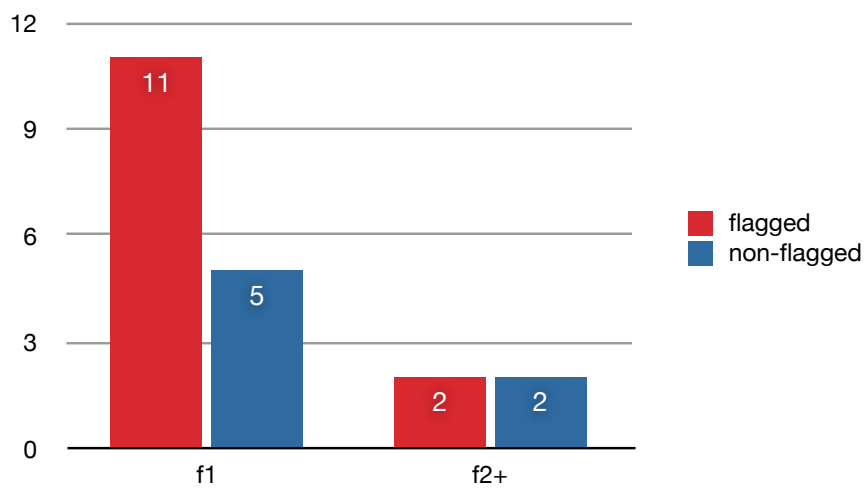


Figure 15. NP flagging and frequency in Milan.

### 6.3.2.2 Types of flagging

The analysis in the previous sections has shown that, at least in Manchester, the insertion of multiword NPs is significantly more flagged than that of lone nouns. In Milan a difference is also observable, but, due to the very low number of tokens, no statistical comparison can be conducted. Besides the quantitative difference, it could be worth asking whether there is also a difference in types of flagging between lone nouns and NPs or, in other words, whether different flagging strategies are used for the two categories.

**Table 4** presents the flagging devices used for NPs in the two groups. There is no statistical difference between NPs and nouns in Manchester with regards to the flagging devices. No difference is found in Milan either, although it should be borne in mind that the low number of NPs and the several empty cells could affect the reliability of statistical tests.

From a purely descriptive point of view, we observe that the flagging strategies used for nouns and NPs are quite similar. Unfilled pauses are a prevalent device in both Manchester and Milan. Lengthening is also quite frequent in Manchester, consistently with the patterns seen for lone nouns. Retracing also appears relatively frequently, once more mirroring the strategies for lone nouns.

The statistical analysis reveals, however, that the difference between nouns and NPs in Manchester with regards to flagging is only quantitative, i.e. it does not involve the type of flagging produced but only its frequency. In Milan no significant difference between lone nouns and NPs, be it quantitative or qualitative, can be observed.

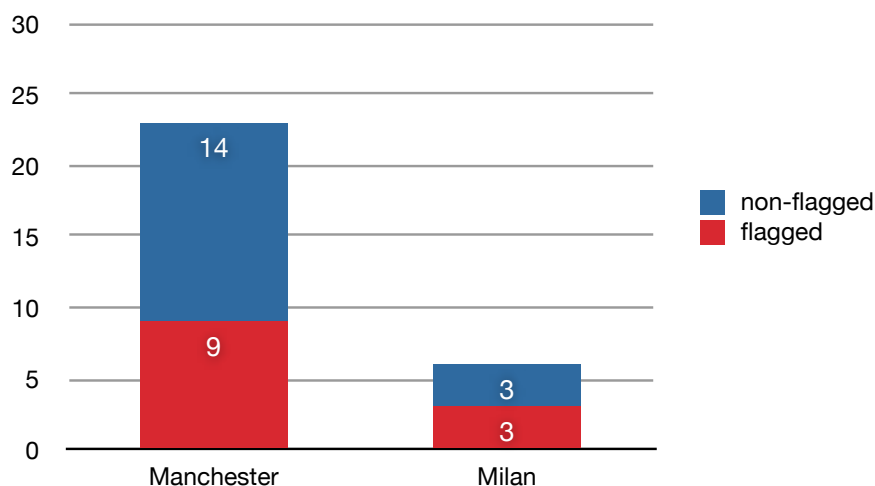
	MANCHESTER	MILAN
length	12	4
LFP	2	0
metalinguistic	3	1
NLFP	3	1
paralinguistic	1	0
retracing	7	1
UP	11	6

**Table 4.** Flagging strategies used for inserted multiword NPs.

### 6.3.3 Flagging of adjectives

After lone nouns and multiword NPs, single adjectives are the following most frequently inserted grammatical category in the data. Overall figures, however, are quite low, with a total of 23 tokens in Manchester and only 6 in Milan.

This is represented graphically in **Figure 16**.



**Figure 16.** Flagging rates for adjective in Manchester and Milan.

Statistical analysis shows that the observable difference between the groups in the rate of flagging for insertions is not significant. Due to their size, the raw numbers do not allow to make any strong claims about the difference between the groups. Still, it is possible to make some observations, especially with regards to the Manchester data.

The overall rate of flagging for adjectives in Milan is higher than the rate for lone nouns, but lower than the one observed for multiword noun phrases.

Also, switched attributive adjectives are virtually absent from the data.

As mentioned earlier in section **6.2.2**, only one adjective in Manchester is unambiguously attributive. The insertion is reproduced in example (62) below.

<b>62.</b> *LUC:	e	lei	<ha	un:> [/]	ha	un	accento
	and	she	have-PRS.3SG	DET	have-PRS.3SG	DET	accent
molto [//]	sì	molto	posh@en	molto: ##	proprio	da	
very	yes	very		very	just	from	
londinese	del	sud	ma	molto: [/]	molto [/]	snob@0	così .
Londoner	of.DET	south	but	very	very		so

*'and she has a... she has an accent very yes very posh very like a Southern Londoner but very very snob like that'*

[Manchester 2]

In the example, the adjective *posh* is placed postnominally, thus following Italian rules. It is also modified by the intensifier *molto* 'very'. We can also note how these lone occurrence of an attributive adjective is flagged in several ways: first by the lengthening on *molto*, followed by some retracing. After the insertion, the speaker engages in a repair sequence by offering an explanation of the adjective *posh* and finally provides a translation equivalent – the noun *snob*, incidentally an established English borrowing in the Italian lexicon used both as a noun and an adjective. The presence of a great amount of flagging in the environment of the only attributive insertion in the data may perhaps be indicative of its low frequency, as well as the fact that the speaker himself feels the insertion of an English attributive in the Italian postnominal position to violate some grammatical constraint; it is equally plausible, however, that the flagging is due to purely discourse-pragmatic reasons, and that the extended repair following the flag is simply a lexical search. Aside from the example reported in (62), all adjectives are predicative.

With regard to the flagging strategies employed by speakers for adjectives, the position and types of flag are not dissimilar from those observed for lone nouns and multiword NPs. A single pre-switch flag is the category with the most frequent position for flags, both in Manchester with six occurrences out of fourteen flags and in Milan, with two out of three flags. The types of flag are distributed amongst the different categories, at least in Manchester, where a sufficient number of tokens is present. The most frequent flagging strategy appears to be lengthening. The flagging types used are reproduced in **Table 5**.

	MANCHESTER	MILAN
length	5	0
LFP	0	1
metalinguistic	2	1
NLFP	1	0
paralinguistic	1	0
retracing	2	0
UP	3	1

**Table 5.** Flagging strategies used for inserted adjectives.

### 6.3.3.1 Flagging and frequency

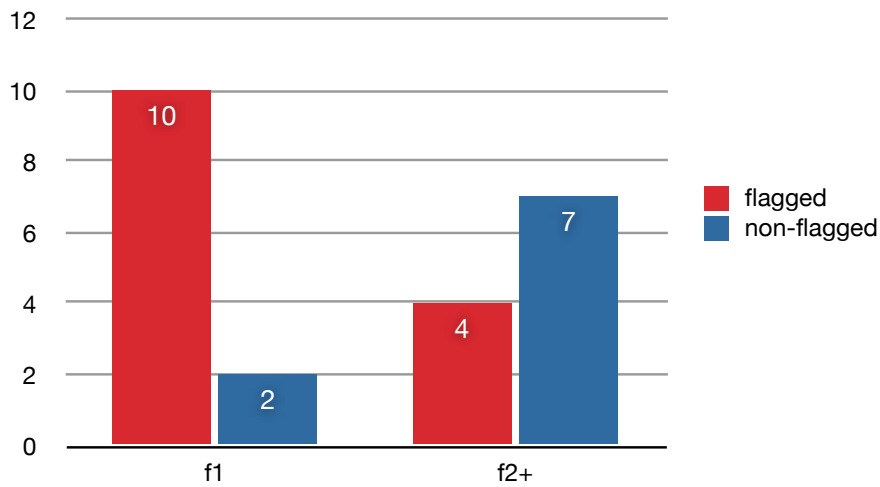
In spite of the numbers being very low, it is still possible, at least for Manchester, to see whether the rates of flagging for adjectives are related to the frequency in the data. In Milan this is not possible, because of there being only six tokens, none of which is of the same type.

Out of the twenty-three adjective tokens in Manchester, twelve only occur once in the data; the remaining eleven tokens are distributed amongst four types.

The adjectives found are *British* (2 tokens); *fictional* (2 tokens); *posh* (3 tokens); *typology-friendly* (4 tokens). It should be noted that while these adjectives are more frequent than others in the data, they are not more widespread, since all the occurrences of each type are found within the same conversation.

The adjective *fictional* (appearing in **Manchester 1**) is only used by one of the speakers. The other three adjectives, on the other hand, are produced by both speakers in the conversation in which they appear. The repetition of the adjective by both speakers, especially when occurring in adjacent turns, can be seen as discourse strategy aimed at establishing inter-speaker cohesion (see Auer, 1998; Gardner-Chloros *et al.*, 2000), thus being unrelated to the actual frequency of the item in discourse at large.

In order to group adjectives according to frequency, the same categories used for nouns and NPs were used, i.e. f1 and f2+ for items with one and more than one occurrence respectively. A graphic representation of the Manchester adjectives and their flagging rates is given in **Figure 17**.



**Figure 17.** Adjective flagging and frequency in Manchester.

In the f1 category, only two adjectives appear with no flagging at all, while the majority – ten out of twelve – is flagged. The picture in the f2 category is quite different, with seven tokens out of eleven being non-flagged and only four flagged. A Fisher’s exact test carried out on the data reveals the difference between the groups to be significant ( $P = 0.0361$ ). It can be argued that at least in Manchester data, a link exists between the presence of flagging and frequency of the adjective.

#### 6.3.4 Flagging of verbs

As discussed in section 6.2.3, insertion of verbs is particularly rare, with only five tokens in Manchester and four in Milan. No verb token is found more than once in the data. These insertions are typically single verb forms, as seen in examples (54) to (59), although there are also a handful of occurrences in which a verb phrase is inserted (i.e. a verb together with a complement NP). This is shown in example (63), in which the speaker inserts the Italian verb *maltrattare* ‘to abuse, ill-treat’, together with its direct object complement *le tate* ‘the nannies’.

The example is reproduced below.

63. \*KEL: that's <because if she loses you> [//] (be)cause she can afford to <uhm@0

maltrattare@it      le@it      tate@it> [>] .  
abuse-INF            DET        nannies

[Milano 4]

It was noted earlier how the inserted verb forms in both data sets tend to be a) non-finite and b) at points of linear equivalence in the two languages. A possible explanation for this lies in the fact that these forms exhibit a higher level of congruence with other-language equivalents than finite forms fully inflected for person and number. It is therefore perhaps surprising to note how all verb insertions in both Manchester and Milan conversations are flagged. Further to this, three of the five Manchester insertions and one of the four Milan insertions are produced with more than one flagging device or, in other words, are quite heavily flagged (remember that one single ante-switch flag is the overall most common flagging strategy for inserted materials in both data sets). With regard to the amount of flagging, verbs therefore seem to pattern differently from other grammatical categories in the data.

## 6.4 Concluding remarks

In this section, the data from both groups have been analysed from a quantitative perspective. Most of the other-language material falls within the category of insertional code-mixing (Muysken, 2000), but other types are also present.

The occurrence of flagging of inserted material in the two corpora as a whole has been presented. The two corpora present similar rates of flagging for inserted other-language material when considered as a whole.

The data have subsequently been analysed according to different grammatical categories. In both Manchester and Milan lone nouns are overwhelmingly the most frequently-inserted category, followed by multiword NPs. In both groups and for both categories, a relationship holds between the frequency of an inserted word in the data and the occurrence of flagging. Less frequent items overall seem to be more flagged than more frequent items, with some differences in the strength of this effect between Manchester and Milan.

A comparison between switched lone nouns and lone nouns in monolingual speech also reveals a significant difference in the occurrence of flagging in the two contexts. Switched lone nouns are twice as likely to be produced with a flag than non-switches.

Adjectives are the category with the next highest number of insertions, even though figures are considerably lower than for nominal constructions. At least in Manchester, a significant effect of frequency on flagging rates can be observed, similarly to nominal constructions. No statistical analysis could be carried out in Milan due to the figures being too low.

A small set of verb insertions are also encountered in both data sets. All of the examples present some degree of flagging.

On the whole, it appears that together with the frequency of an inserted item, its morphological category may also play a part in its appearance with or without flagging. This is supported by the fact that different categories exhibit very different flagging rates in the data. Inserted verbs have the overall highest flagging rate, followed by multiword NPs, adjectives and lone nouns. ¶



# QUALITATIVE ANALYSIS: A CONVERSATION ANALYTIC PERSPECTIVE

In the present chapter, the data are analysed qualitatively through an adaptation of Conversation Analysis methodology applied to CS (see section **2.3.2**). Particular focus will be on the issue of conversational repair (Schegloff *et al.*, 1977) in conjunction with instances of language alternation. By looking at the different methods that speakers adopt to repair the problems – particularly word searches – that arise in speech, and how they respond to their own solutions, it is possible to discover how speakers mark the boundaries of the linguistic code (medium) they use in an interactional episode, and how these boundaries may not correspond to the boundaries between two (or more) languages involved in the conversation (see sections **2.3.2** and **2.4**). Flagging strategies are instrumental in reconstructing participants' orientation to the codes they use in conversation, since they overwhelmingly mark repair, as will be shown.

A brief discussion on the limits of clause-by-clause analysis, which was the approach in the previous chapter, will follow below. The discussion is accompanied by a slight alteration of our working definition of flagging. Subsequently the phenomenon of repair (see section **3.3**) will be illustrated with examples from the data, as an application of an interpretive perspective to the study language alternation.

As shown in the previous chapter, speakers in Manchester and Milan largely appear to be using the same flagging strategies in conversation. Examples are therefore drawn from both groups, with no marked contrasting of group-specific patterns.



1572 \*KEL: anyway # uhm:@0 [>] # but now it's really strange  
1573 because # he was gone studying English in: London  
1574 for: two months or whatever it was .  
1575 \*EMM: &=laughs [<] .

[Milano 2]

In line 1568 speaker Kelly utters the Italian word *crisi* 'crisis'. It is part of a mixed noun phrase, *relationship crisi*. A medium-length UP before the noun phrase interrupts the flow of speech, thus drawing attention to (flagging) the following item. In the quantitative clause-by-clause analysis the pause preceding the NP would simply have been registered as the only flagging for the switch.

If we look at the insertion in its wider environment, however, it becomes apparent that more flagging activity (here in the form of an explicit metalinguistic commentary) is present. The speaker engages in a rather elaborate sequence in which she accounts for her behaviour, namely the use of a particular Italian word. The word *crisi* is used in this context because, the speaker tells her interlocutor, it is what everyone says (lines 1569-1571). *Everyone* presumably refers to Italians, who routinely use this word to describe a couple's relationship problems, but it could also be other English L1 speakers in Milan. The subsequent quotation in line 1570 *they've had their crisi* is in fact a calque of phrases commonly used in Italian to refer to couples experiencing relationship problems (*sono in crisi/ hanno avuto una crisi*, lit. 'they're in crisis/they've had a crisis').

While only little flagging can be seen in the same clause as the switch (in the form of an UP) more of the same phenomenon can be found if the angle changes, and one looks at larger data fragments, rather than isolated clauses. If flagging only occurred outside the clause in which the switch occurs, therefore not adjacently to it, there are some instances of CS which would have simply been classified as non-flagged in the quantitative analysis, even though there is flagging clearly referring to them in the immediately preceding or following turns. In the quantitative clause-by-clause analysis in the previous chapter, the pause preceding the NP *relationship crisi* would simply have been registered as the only flagging for the switch.

As shown in the example in (64), the definition of flagging adopted in the present study, which includes the feature of adjacency of the flag to the switch (see section 5.1 for the definition), will have constrained the analysis. While the overwhelming majority of flags encountered in the present data are indeed adjacent to the switch, there are cases in which flagging relating to a switch can occur in a different clause or even in a different turn from the switch that it refers to. This only seems to occur with the more explicit flags, such as pre-empting, justification and translation equivalents, i.e. those for which an exclusively structural definition is not possible.

Whereas the previous analysis was concerned with providing an overview of the quantity and type of flagging in the two groups of speakers, the focus is now on participants' own interpretation of language alternation, as highlighted by the flagging they produce in interaction, and the procedures they use to repair problems that arise in the unfolding of conversation. Such organisational issues are typically the subject matter of CA, which is the perspective adopted in the present section. The application of the methods of CA implies that a strict clause-by-clause analysis will be replaced by an analysis of sequences possibly stretching over several turns. This of course involves a shift from a more structural view of language alternation to a more markedly interpretive approach. The definition of flagging, while including the same inventory of discourse phenomena presented in the methodology (chapter 5), will also have to be adapted for the present analysis. In particular, since the unit of analysis is no longer the clause, we will expect flagging in the vicinity of the switch, rather than only adjacently to it. The example in (64) already showed how this may yield different results and reveal additional information about the status of inserted items in conversation.

## 7.2 Applying Conversation Analysis to the data: looking at repair

As already seen in the quantitative analysis in chapter 5, instances of CS have a relatively low frequency in the present data. Because of this, the signalling value of these occurrences, that is, the effect of “otherness” that they create when inserted in speech is likely to be relatively high. Auer (1999) notes that the more frequently language alternation occurs, the less likely single other-language items are to be meaningful *qua* other language. When the presence of other-language elements in speech becomes more frequent and widespread, single instances of language alternation may gradually lose their discourse or pragmatic value. In other words, with increased presence these elements become integrated in the base and can come to be seen as part of the base code or “medium” of the conversation, losing all pragmatic/stylistic value as contextualisation cues. In the present groups of speakers, however, this loss of pragmatic/discourse value does not seem to have occurred. It may provisionally be assumed, that at least some other-language items inserted in speech are likely to be rather interactionally salient, and be treated as deviant by speakers themselves.

If we focus on single cases of language alternation, leaving established borrowing aside (i.e. words that in the transcription phase have been assumed to belong to both English and Italian), one issue is how to distinguish cases that are part of the current medium of interaction (from a participants’ perspective) from those which are not (see sections 2.3.2.5 and 2.4.3 for the definition of medium). In other words, some instances of what externally looks like language alternation may be part of what participants see as the medium, while other cases may be genuine cases of departure from current medium, which are perceived as such by speakers themselves. The language of an item is not a sufficient criterion, since elements from more than one language may be part of the current medium, as argued in section 2.4.3.

In order to uncover the underlying orderliness of interaction between bilinguals, and the boundaries of the medium that they are using, it can be useful to look at cases where speakers encounter specific problems, the methods they use to solve them and, crucially, how they orient to their own solutions. In CA, these methods (or procedures) used by speakers are collectively known as “repair”

(see Schegloff *et al.*, 1977). As argued in section 3.3, there is a substantial overlap between repair and the discourse phenomena falling under the category of flagging. Similarly to flagging, repair is seen as an interruption of the speech flow (see Fox *et al.*, 1996; Rieger, 2003; Liddicoat, 2007: 177 for a detailed review of repair). While the concept of repair focuses more on the process of dealing with a trouble occurring in talk, flagging can be the overt manifestation of this process. Amongst the problems that can trigger repair a common one is that of retrieving the “next item due” (Fox *et al.*, 1996) or, in other words, a lexical search.

An example from the data is provided in (65):

**65.** *Kylie and Amiria are two friends talking about their children, who are both in primary school. They are expressing concern about the children’s schoolbags being too heavy.*

802 \*KYL: yeah but it’s better than our day where we just  
803 had <the: # &s> [/] # a sling over your shoulder bag with  
804 all the same am(ount) [/] amount of things .  
805 \*KYL: at least nowadays you got rucksacks to use .  
806 \*AMI: yeah they’re just gigantic anyway .

[Milano 3]

In line 803, we observe an interruption of the speech flow in conjunction with what appears to be a lexical search. The determiner *the* is lengthened, is followed by a pause and by an attempt to utter the next item due (&s) which is immediately cut off and followed by a further pause, before the determiner is retraced and *a sling* is inserted as the solution to the problem. After the solution has been uttered, Kylie immediately resumes her turn, with no further orientation to the incident by either speaker. Liddicoat (2007) defines cases such as this as same-turn self-initiated self-repair, also noticing how repairs tend to occur preferentially close to the trouble source (supporting the findings of the classic study of repair by Schegloff *et al.*, 1977).

In bilingual conversation the deployed repair may be an instance of language alternation, if the solution to the problem is from a different language to that used in the immediately preceding context. Gafaranga (2000) has looked at repair strategies in bilingual conversation as a way of discovering the medium used by

participants. In particular, he focuses on what he calls the “search for the *mot juste*” (2000: 347), or apt expression. Switching for the *mot juste* is a well-known phenomenon in studies of bilingual conversation (see Poplack, 1988; Jones, 2005). Gafaranga argues that bilingual speakers adopt one of two strategies to tackle this problem: they will either solve it with an element from the current medium of interaction or they will use an element whose occurrence will in turn require further repair. In the latter case, speakers are deviating from the medium. Crucially, because the medium is not a language, the *mot juste* may or may not look like a switch to the external observer. Gafaranga (2000) speaks of same-language and other-language repair when the *mot juste* is interpreted by participants as part of the medium, while he calls medium-repair the instances of departure from the medium of the conversation. All these types of occurrence will be discussed in the present chapter. The focus in the analysis will mainly be on the different repair strategies utilised by speakers in the data with reference to word searches, and how these intersect with language alternation and flagging. Focus will be on the role flagging plays in reconstructing participants’ interpretation of the interaction. It will be argued that flagging can contribute to highlighting the boundaries of the medium from the perspective of interactants. The different ways in which seemingly similar instances of language alternation are oriented to by speakers can reveal which items are treated as normative, thus being part of the medium and which ones are not, therefore falling outside the boundaries of the medium, which are negotiated by speakers. The observation of repair in conjunction with lexical searches can enable the analyst to discover the medium adopted by speakers, by looking at how the different solutions to the problems encountered in talk are deployed and ratified.

### 7.2.1 Repair in monolingual contexts

To illustrate the features of repair in conjunction with lexical searches, let us first look at two monolingual extracts taken from the corpus. The first example in (66) comes from English L1 speakers in Milan:

66. *Kelly and Emma are friends and former colleagues talking in their formerly shared workspace* (see example 64).

1065 \*KEL: &ab [//] ok@0 we went to Mexico because I went in  
1066 <San\_Diego@0 and it's only twenty minutes> [>1] from  
1067 Tijuana@0 [>2] .  
1068 \*EMM: <oh@0 yes mmhm@0> [<1] .  
1069 \*EMM: yeah [<2] .  
1070 \*KEL: and so one day we decided we'd walk across the  
1071 border .  
1072 \*EMM: mmhm@0 .  
1073 \*KEL: and uhm@0 you know how they've got like the: +...  
1074 \*KEL: well just like here you've got the **tourist traps** .  
1075  
1076 \*KEL: well they had # a donkey painted like a zebra  
1077 for the tourists to take # pictures with [>] .  
1078 \*EMM: oh@0 [<] my word [>] !  
1079 \*KEL: and: [<] she goes +"/.  
1080 \*KEL: +" oh@0 my gosh # how did they get zebras in  
1081 Mexico ?  
1082 \*KEL: +" I didn't know zebras were in Mexico !  
1083 \*KEL: and # I mean it was really obvious that it was  
1084 a donkey painted [>] .

[Milano 2]

In the above extract, Kelly is telling her friend Emma about a trip to Mexico with a friend. In line 1073, she suddenly encounters a problem, in that she is lacking the exact word (or phrase) she would like to use at that particular point in speech. After a hesitation marker (NLFP *uhm*), she starts to retrieve the next item due, but cannot seem to find it and trails off (*you've got like the: +...*). She eventually retrieves the phrase she had been looking for (*tourist traps*) and inserts it after reiterating (retracing) her previous utterance in a slightly altered way (line 1074). As soon as she has deployed the solution to the problem that had arisen, she simply continues with her narration sequence. Her interlocutor does not seem (at least audibly) to regard the phrase as problematic or to require further explanations, and the conversation continues with no further interruptions.



If we look at the passage from the point of view of the hesitation strategies used we notice that the problem is signalled quite early by the NLFP *uhm* and then again by the use of the filler *like* and the lengthening of the determiner *the* at the point where the next lexical item would be expected. The fact that the speaker then retraces her utterance to insert the phrase *tourist traps* corroborates the interpretation that a word search problem had indeed been encountered, and has subsequently been repaired by the speaker. After the incident has been repaired, however, no other flagging devices are produced, and Kelly's narration resumes where it had been interrupted.

The example in (67) below contains two occurrences of a very similar problem close to one another, this time in a conversation between Italian L1 speakers in Manchester:

**67.** *Luca and Paolo are friends. Paolo is telling Luca about some recent career advice he has received.*

509 \*LUC: quindi potre(sti) [///] lei [///] Hane@0 cos' è che ti  
 510 ha detto ?  
 511 \*PAO: Hane@0 mi ha detto in pratica che # e: <devo> [//] uh@0  
 512 posso provare a mandare a delle ehm ## e: come  
 513 si dice delle: ## **degli articoli** ## giocando sui  
 514 problemi sociali socioculturali che possono: # venir  
 515 fuori dalla traduzione: eh # <a questi:> [/] <a  
 516 questi:> [//] a queste **riviste** .  
 517 \*PAO: New\_Voices@0 # conosci ?  
 518 \*LUC: no@0 .  
 519 \*PAO: no .

---

509 \*LUC: *so you could [///] she [///] Hane@0 what's it that she*  
 510 *has told you ?*  
 511 \*PAO: *Hane@0 basically told me that # uhm: <I have to> [//] uh@0*  
 512 *I can try to submit to some uhm ## uh how*  
 513 *do you say some: ## **some papers** ## that play on*  
 514 *social sociocultural problems which may: # arise*  
 515 *from the translation: uh # <to these:> [/] <to*

516            *to these:> [//] to these journals .*  
517 \*PAO: *New\_Voices@0 # do you know that ?*  
518 \*LUC: *no@0 .*  
519 \*PAO: *no .*

[Manchester 2]

In the example, speaker Paolo is faced with two lexical searches. In lines 512-513 he signals the problem by the use of a hesitation marker (NLFP *ehm*), an unfilled pause and by an explicit comment that he is searching for a word (*come si dice*, ‘how do you say’). The latter could represent a form of forewarning for what is about to come, as well as a request for collaboration from the interlocutor. Paolo continues by retracing the determiner *delle* ‘some’ (fem. plur.), and inserting a further unfilled pause in line 513. He finally retrieves the desired word *articoli*, ‘papers’ which he inserts by further retracing and modifying the determiner *degli* ‘some’ (masc. plur.) (notice the change of gender from feminine *delle* ‘some’ (fem.plur.), line 513). Once the problem has been solved, it appears the solution deployed by Paolo has been accepted, since neither he nor his interlocutor Luca make any further comments.

A very similar problem is encountered immediately afterwards (lines 515-516). Paolo interrupts the speech flow again. He signals he is going through a word search by retracing the prepositional phrase *a questi*, ‘to these’ (masc.) together with some vowel lengthening. Once more, when the word is found the phrase is retraced one last time, with the gender of the demonstrative *queste* ‘these’ (fem.) made to agree with the noun *riviste* ‘journals’. The interaction then continues with no further reference to the repair episode.

While the subsequent analysis will focus on bilingual conversation, examples (66) and (67) have shown that a similar problem, i.e. that of word searches, is also found (and addressed) in monolingual conversation. The problem therefore occurs not only in bilingual contexts, but is a consequence of the basic architecture of speech, which implies that obstacles encountered in production will have to be solved on the spot (see Gafaranga, 2000: 330). It is therefore reasonable to expect that the issue will be found in monolingual and bilingual language alike.

In bilingual speech, repair in conjunction with a lexical search can be of particular interest within a CA framework, in that it represents an incident in which the norms which speakers share with reference to language choice may be tested and revealed. By looking at the ways in which speakers signal and solve the problems found in speech production, analysts may attempt to reconstruct the scheme of interpretation used by participants in a given episode.

These problem sites are then privileged objects of investigation for analysts wanting to uncover the boundaries of the conversation's medium, since it is in problem sites, with an apparent violation of a norm, that the underlying rules shared by participants become visible.

In the above examples, several instances of flagging ("trouble markers" to quote Gafaranga, 2000: 333) signal to the interlocutor that the speaker is dealing with a trouble source (a word search) and that he has not terminated his turn. Theoretically, they also allow the interlocutor to intervene to solve the problem (see example **75** in section **6.3** below), even though it seems to be more common for the current speaker to solve the problem herself (Schegloff *et al.*, 1977; Liddicoat, 2007). As soon as the search is finished, however, no more repair activity is found, and the speech flow resumes. It can be inferred that for those cases the type of repair deployed (i.e. the word inserted) is suitable, and does not constitute an instance of deviance from the medium of the conversation. Because for the external observer the passage is monolingual, however, this observation may seem rather inconsequential.

### 7.2.2 Repair in bilingual contexts

Now that we have seen some instantiations of repair (for a word search) in a monolingual context through the use of different discourse phenomena, our attention will shift to bilingual contexts. Before focussing on cases of repair involving language alternation, however, we shall examine a bilingual case which involves no repair. The example, presented in (**68**), comes from the Milan data.



‘Year One in primary school’ to the *quinta* ‘Year Five’. Speaker Kylie then says of the school that *it goes all the way*, meaning that it continues beyond the primary level. Amiria checks her own understanding of this in line 254 by glossing Kylie’s preceding turn with *media e liceo*<sup>34</sup> ‘middle and secondary school’.

What is of interest here is the fact that none of the Italian elements inserted by Amiria are marked by flagging or other discourse phenomena. There are no instances of pauses, hesitations or other flags that make the inserted elements stand out from the surrounding discourse. The words are not inserted as a solution to a problem that has arisen in speech (e.g. the unavailability of a specific lexical item). Amiria’s interlocutor Kylie also does not seem to pick up on the presence of words that appear ‘other-language’ to the external observer. Looking at the example within the framework of the medium, it is possible to infer that the particular words used by Amiria are felt by both speakers to be part of the medium of the conversation, since they are not oriented to by either as being deviant, dispreferred or requiring a justification. Equally importantly, it should be noted that language alternation does not occur exclusively as part of a repair sequence, but can happen with no perceptible interruption of the speech flow. The example in (68) contrasts with the ones presented in (69) from Manchester and (70) from Milan below, in which the instance of language alternation is preceded by flagging. In these examples the delivery of a speaker’s turn is disturbed by the occurrence of a problem, which is then repaired.

**68.** *Two friends, Sara and Antonio are having coffee in Sara’s house. Sara is talking about a common friend and an invitation that she recently received from a lecturer at her university.*

874 \*SAR: e allora gli ho detto che appunto <ho inv(itato)>  
875        [//] avevo ricevuto questo invito da Catherine@0 .  
876 \*SAR: Catherine@0 <era la:> [/] era la supervisor@en di: ehm  
877        Paul@0 [>] .  
878 \*ANT: ah@0 [<] sì sì .  
879 \*SAR: mmhm@0.  
880 \*SAR: che abita qua intorno [>] .  
881 \*ANT: mmhm@0 [<] .

---

**34** To be exact, Italian *liceo* in Italian denotes a specific type of *scuola superiore* ‘secondary school’.

---

874 \*SAR: *and so I told him that exactly <I have inv(ited)>*  
 875 *[//] had received this invitation from Catherine@0.*  
 876 \*SAR: *Catherine@0 <was the:> [//] was the **supervisor**@en of: uhm*  
 877 *Paul@0 [>].*  
 878 \*ANT: *ah@0 [<] yes yes.*  
 879 \*SAR: *mmhm@0.*  
 880 \*SAR: *who lives around here [>].*  
 881 \*ANT: *mmhm@0 [<].*

[Manchester 1]

In the example, Sara encounters a problem while formulating her turn, as signalled by the lengthening of the determiner *la* ‘the’ (fem. sing.) and the subsequent retracing of the verb and determiner *era la* ‘was the’ (line 876). Once she repairs the problem by uttering *supervisor*, however, the conversation carries on as normal. Neither Sara nor her interlocutor can be seen to treat the insertion of *supervisor* as a deviant element requiring further repair. Sara repairs the problem that she encounters by resorting to another language from the one she was using. Once she has produced the word *supervisor*, however, neither she nor her interlocutor Antonio can be shown to orient to the “otherness” of the English element. It could be argued, therefore, that *supervisor* is a suitable solution to the lexical search initiated by Sara, and that her interlocutor does not request further explanation (i.e. he accepts the solution). *Supervisor*, therefore, is still perceived as part of the current medium.

In (70) an example from Milan speakers involving a similar type of repair is presented.

**70.** *The conversation takes place in an office on a Friday afternoon. Kelly is talking to Emma about a friend’s partner.*

1503 \*KEL: *or my friend Sophie@0 # it’s another gossip story .*  
 1504 \*EMM: *what’s happened to her ?*  
 1505 \*KEL: *well she’s been with the same guy for like I*  
 1506 *guess seven years now .*  
 1507 \*KEL: *and the +...*

1508 \*KEL: oh@0 Sandro@0 is such a nice boy .  
 1509 \*EMM:mmhm@0 [>] .  
 1510 \*KEL: <he's such a> [<] [//] # he's <just a> [//] such a  
 1511 little **cucciolo**@it .  
    puppy  
 1512 \*KEL: he's just so perfect .  
 1513 \*KEL: he cooks .  
 1514 \*KEL: then he cleans .  
 1515 \*KEL: and he does all of the &l +...  
 1516 \*KEL: the only thing that he doesn't do is ironing .  
 1517 \*EMM:ah@0 .

[Milano 2]

Kelly has been talking to Emma about a friend's relationship issues. Now she introduces *another gossip story*. She starts by describing her friend's partner (lines 1505-1508) but encounters a problem when she tries to characterise him in more detail (line 1510). It appears likely that the word she would like to use (the next item due) is not immediately available to her. The problem sequence starts as an expansion of line 1508 (*Sandro is such a nice boy*). After several false starts and a pause in line 1510, Kelly eventually deploys the Italian word *cucciolo* 'puppy'. Once the word has been uttered, however, neither Kelly or her interlocutor can be seen to pick up on to the insertion, and the conversation continues with no further reference to the incident. Similarly to example (69), the speaker signals a trouble through various flags, and the solution to the problem involves departing from the language being used. This way of solving the issue, however, is not seen as constituting a problem by either participant with regard to the medium. The examples seen so far have shown two different methods of solving problems caused by word searches. The difference between the two, however, may only be appreciated by the external observer. In examples (66) and (67) the problem was resolved by the current speaker by using the current language of interaction. In examples (69) and (70) on the other hand, the speaker resorted to the other language. If we examine the type of flagging which speakers use, however, we find little or no difference between cases where the repair is deployed in the same language or in a different language from the surrounding discourse. Speakers'

orientation to the instances of repair remains the same irrespective of the language of the repair element, and the methods speakers use to tackle the problems are very similar in all cases. In no case does further flagging appear after the solution to the word search has been deployed, and the conversation resumes where its flow had been interrupted by the onset of the problem. In (68), by contrast, language alternation does not occur as a repair to a trouble but happens smoothly, with no prior flagging, or requests of an account by the interlocutor.

In all these cases then, it is possible to argue that speakers are attending to what they see as the current medium of conversation. In (68) other-language insertions are not oriented to as problematic by speakers; they are not part of a repair sequence and are not picked up on by speakers once they have been produced. In (69) and (70), similarly, the solutions that speakers propose to the problems encountered in conversation are interpreted by them as complying with their current norms. With reference to these examples, it would be possible to say that speakers have agreed on a bilingual medium. From an external observer's perspective we could qualify this medium as basically one language with some elements from the other.

### 7.2.3 **Deviating from the medium**

The examples below illustrates a different type of repair, and a different way in which flagging strategies are employed by speakers. In the previous examples the interaction continues regularly immediately after the insertion of the solution to the word search problem. In the following examples, on the other hand, the solution adopted seems to require further repair work by the speaker. This is achieved by additional, different types of flagging.

The first extract is the same as that seen in example (64), reproduced here as example (71) for convenience:





sequence, in which she explains that *crisi* is the word everyone uses. She continues by reporting the entire stock phrase as a direct quotation (lines 1570-1571). Interestingly, she does not quote the expression in Italian, but translates it into English whilst preserving the Italian *crisi*. Only after justifying the use of that particular noun, does she bring the sequence to a close in the absence of explicit feedback from Emma (*anyway*, line 1572) and then resumes her narrative. Throughout her repair work, the speaker keeps the Italian version of the word, which is perceived as not being part of the medium, rather than producing a translation. At the same time, however, she feels she needs to provide her interlocutor with an account of her language choice, thus showing that she sees it as deviant. The account comes in the form of explicit justification, which represents an instance of flagging.

The following example, reported in (72), comes from Manchester. It presents a very similar problem to that seen in example (71).

**72.** *Paolo is talking to his friend Luca about a novel he has read in both English and Italian and how some elements get lost in the translation from the English original into Italian.*

164 \*PAO: e quindi c' è questa perdita .  
 165 \*PAO: poi ci sono per esempio altre: uhm@0 # perdite secondo me .  
 166 \*PAO: sono le: uhm@0 e ## quota(tions)@en [///] e: le: citazioni #  
 167 parodiche .  
 168 \*PAO: tipo il personaggio # si ubriaca .  
 169 \*PAO: e inizia +//.  
 170 \*PAO: siccome lui è un amante della letteratura inglese eh # inizia a  
 171 prendere # dei versi dei poeti # e a stravolgerli .

---

164 \*PAO: *and so you have this loss.*  
 165 \*PAO: *then there are for example other: uhm@0 # losses in my opinion.*  
 166 \*PAO: *it's the uhm@0 uhe ## quota(tions)@en [///] uh: the: quotations #*  
 167 *parodying [quotations].*  
 168 \*PAO: *like the character # gets drunk.*  
 169 \*PAO: *and starts +//.*  
 170 \*PAO: *because he is a lover of English literature you know # he begins to*  
 171 *take # some lines from poets # and turns them on their heads.*

[Manchester 2]

Paolo encounters a problem (line 166) as signalled by a series of flags: the lengthening of the determiner *le* 'the' fem. plur, the NLFP *uhm*, the Italian filler *e* and an UP. This long series of hesitation markers seems to come to a close with the insertion of the word *quota(tions)*<sup>35</sup> (line 166), which represents the solution to the lexical search. However, no sooner has Paolo started uttering the word *quotations* that he trails off. He produces some more hesitation markers in the form of Italian filler *e* (with lengthening) and a prolonged Italian determiner *le* 'the' (fem.plur), effectively resuming his search. He then inserts the Italian *citazioni*, the exact translation equivalent of *quotations*. After a brief pause he qualifies the *citazioni* as *parodiche* 'parodying', and then continues with his turn by giving an illustrative example of what he meant.

In the extract Paolo encounters a problem, and after some hesitation he attempts a repair. As he is in the process of doing so, however, he picks up on the solution not being appropriate in the particular context where it has occurred, as signalled by the fact that the proposed solution is not uttered fully but interrupted part-way through the word. He then immediately engages in further repair work in order to provide a solution to the original lexical search problem. The occurrence of the word *quotations* therefore is not seen by Paolo as being part of the medium of the conversation, and has to be repaired.

In examples (71) and (72), speakers have shown a rather different orientation to the solutions they found to the lexical search problems compared with (69) and (70). The use of flagging to interrupt the delivery of talk is similar in all cases, up until the insertion of the solution. After the solution is inserted, however, speakers can be seen to carry out further repair work, or to offer an account of why they used that particular element. The accounts are themselves characterised by flagging (retracing, pauses, hesitations), which again contribute to signal the problematic nature of the repair. In other words, when a problem occurs, it triggers a lexical search. The solution to the problem is eventually uttered, but then the solution itself needs to be rectified. This can be interpreted as evidence that the first

---

<sup>35</sup> In the audio recording the speaker interrupts the delivery of the word. However it is possible to hear the segment [kwəʊ'teɪ]. It was therefore inferred as the most plausible hypothesis, that *quotations* was the word the speaker was uttering.

solutions deployed by speakers are instances of deviance from the current medium, because speakers seek to redress them immediately after they have been uttered by carrying out further repair work.

Instances of deviance from the medium, however, need not necessarily follow the structure seen so far. The example in (73) from Italian L1 speakers in Manchester, shows a similar problem but in a slightly different context:

73. *Sara is talking to Antonio about a writer whose work she has analysed as part of her course.*

1818 \*SAR: è uno scrittore: uno dei migliori insomma nel  
1819 panorama canadese non solo italo-canadese [>] .  
1820 \*ANT: quindi [<] <per avere eh> [>] +...  
1821 \*SAR: <ed è uno> [<] che non vuole essere considerato  
1822 **ethnic**@en # uno scrittore etnico [>] .  
1823 \*ANT: ah@0 [<] .  
1824 \*SAR: vuole essere considerato uno scrittore canadese .  
1825 \*ANT: eh@0 .  
1826 \*SAR: <ovviamente ehm> [>] +...

---

1818 \*SAR: *he's a writer: one of the best ones well*  
1819 *not just in the Italian-Canadian but on the Canadian scene [>].*  
1820 \*ANT: *so [<] <in order to have eh> [>] +...*  
1821 \*SAR: *<and he's one> [<] who doesn't want to be considered*  
1822 **ethnic**@en # *an ethnic writer [>].*  
1823 \*ANT: *ah@0 [<].*  
1824 \*SAR: *he wants to be considered as Canadian writer.*  
1825 \*ANT: *eh.*  
1826 \*SAR: *<obviously ehm> [>] +...*

[Manchester 1]

Sara uses the adjective *ethnic* (line 1822) to describe the writer she is talking about. Unlike in the previous examples, the other-language word is not flagged before being produced: there are no interruptions in the speech flow, hesitations or false starts before the word is uttered. The insertion is completely smooth, and is not signalled by pre-flagging indicating trouble, similarly to example (68). *Ethnic* would at this stage appear to be part of the medium of the conversation.

In fact, the word is not felt to be appropriate by Sara in the current medium. After a brief unfilled pause (a post-flag), she starts a self-repair sequence in which she produces the Italian translation equivalent of *ethnic* in the paraphrase *uno scrittore etnico* ‘an ethnic writer’. She then expands further the meaning of her repair in the subsequent turn *vuole essere considerato uno scrittore canadese* ‘he wants to be considered as a Canadian writer’ (i.e. as opposed to an ethnic one). Throughout the sequence Sara’s interlocutor Antonio only seems to give minimum feedback and has no role in initiating or contributing to the repair sequence. From Sara’s orientation to her own other-language insertion, it may be argued that an element that is not part of the medium has been used. The speaker felt the element was not acceptable and started a repair sequence by providing a correction deemed suitable. This again contrasts with what was seen in examples (69) and (70), where the other-language items were inserted at the end of a lexical search marked by flagging, but were then no longer dealt with by participants. The example in (73) also differs from (71) and (72), in that no flagging is present before the other-language word. The signalling is only found after the word *ethnic*, which may be seen almost as a slip of the tongue, rather than the genuine solution to a lexical search problem. Even in a context characterised by the absence of flagging before the switch, speakers show their sensitivity to the orderliness of the interaction, by immediately flagging and repairing an element perceived as deviant from the current medium.

### 7.3 Shifting the boundaries of the medium

Examples so far have shown that where speakers encounter a problem in the delivery of their speech, they initiate and accomplish a sequence of repair with no necessary input from their interlocutor. It seems indeed to be the case that self-repair, rather than other-initiated repair, is the preferred strategy adopted by speakers when such problems arise (Schegloff *et al.*, 1977). There may be cases, however, where the repair to a problem is accomplished, initiated or requested by the interlocutor, as in the example (74) below. This example also shows that the boundaries of the medium, while largely shared, may not be identical for both speakers at all times and may require some negotiation.

74. Sara and Antonio are talking about buying houses. Antonio talks about his plans and describes what he is looking for in a property.

- 1666 \*SAR: e tu # hai pensato alla casa ?  
1667 \*SAR: ci stai pensando [>] ?  
1668 \*ANT: l' [<] anno prossimo # compero .  
1669 \*SAR: sì .  
1670 \*ANT: comprerò: &=coughs .  
1671 \*ANT: beh dipende # quello che trovo .  
1672 \*SAR: mmhm@0 .  
1673 \*ANT: **three@en bedroom@en house@en** .  
(1.0)  
1674 \*SAR: cosa ?  
1675 \*ANT: **three@en bedroom@en house@en** .  
(1.0)  
1676 \*SAR: <**three@en** [///] ah@0 vuoi> [>] fare il **three@en bedrooms@en** ?  
1677 \*ANT: <con un> [<] +...  
1678 \*ANT: mmhm@0 .  
1679 \*SAR: mmhm@0 .  
1680 \*ANT: con il giardino davanti e dietro .

- 
- 1666 \*SAR: *and you # have you thought about a house?*  
1667 \*SAR: *are you thinking about it [>]?*  
1668 \*ANT: *next [<] year # I'm going to buy.*  
1669 \*SAR: *yes.*  
1670 \*ANT: *I'll buy: &=coughs.*  
1671 \*ANT: *well it depends # what I find.*  
1672 \*SAR: *mmhm@0.*  
1673 \*ANT: ***three@en bedroom@en house@en.***  
(1.0)  
1674 \*SAR: *what?*  
1675 \*ANT: ***three@en bedroom@en house@en.***  
(1.0)  
1676 \*SAR: <***three@en*** [///] ah@0 you want to> [>] go for ***three@en bedrooms@en?***  
1677 \*ANT: <*with a*> [<] +...  
1678 \*ANT: *mmhm@0.*  
1679 \*SAR: *mmhm@0.*  
1680 \*ANT: *with a front and a back garden.*

[Manchester 1]

In line 1673, Antonio uses the phrase *three-bedroom house*. The phrase is not flagged, and Antonio seems to have terminated his turn. After a one-second silence<sup>36</sup>, Sara asks for a clarification/repetition (line 1674), thus requesting a repair (because she could either not hear or not understand what Antonio had said). Antonio repeats the phrase with no changes (line 1675) and his intervention is followed by further silence from Sara. She eventually shows her understanding after a false start in line 1676. Since the false start seems to be a repetition of the phrase which had been the object of repair, it may be argued that Sara had not immediately grasped its meaning and is about to ask for a further clarification.

The absence of flagging in this extract is particularly revealing from the perspective of the medium. The lack of any flagging in the speech of Antonio signals that he does not see his own other-language insertion as deviant, but rather as part of the medium. Sara's silence and subsequent request for repair, however, show lack of understanding on her behalf. The fact that the sequence is repeated verbatim by Antonio, rather than in translation or with an explanation, reinforces the idea that he does not see his other-language insertion as problematic in this context; if he had, he would have probably responded to Sara's request with a translation (reiteration for clarification in the other language is a common function of language alternation. See Gumperz, 1982: 78; Gardner-Chloros *et al.* 2000; Alfonzetti, 1998). The fact that he simply repeats his turn is quite significant as it reveals his orientation. After a further silence, Sara eventually ratifies the other-language phrase by using it herself. It may well be the case that *three bedroom house* is perceived by Sarah as deviant, but that she then decides to comply with Antonio's choice. As noted by Auer (1998: 7), repetition of the same word in the same language can also be seen as a "strategy of establishing textual coherence". The medium of a conversation is being constructed in interaction, when speakers can be shown to re-adjust the boundaries of their own categories.

A similar case, where the exact boundaries of the medium do not seem to coincide for speakers, is presented in (75):

---

<sup>36</sup> Pauses between turns are not marked in the transcription. They have been included in the presentation of this example in order to make understanding of the example easier.

75. *Friends* Paolo and Luca are in a university café talking about translation and adaptation problems in books and TV series. Paolo then introduces a specific example.

435 \*PAO: per esempio tornando alla Tata c'è una scena # in  
436 cui ehm c'è la [///] il compleanno della madre .  
437 \*PAO: e [/] e praticamente nella versione italiana .  
438 \*PAO: cioè quella puntata non ha molto senso perché  
439 fa vedere # lei che prepara i festoni: uh@0 +"/.  
440 \*PAO: +" happy@en birthday@en mama@en .  
441 \*PAO: e arriva la zia .  
442 \*PAO: e [/] e nella versione italiana c'è un uhm@0 ## voiceover@en .  
443 \*PAO: come si chiama .  
444 \*PAO: mo' non mi viene in italiano il: [>] .  
445 \*LUC: sì [<] .  
446 \*PAO: una [/] una voce: fuori campo [>] # che [/] <che  
447 dic(e)> [//] de(l)la [/] della Tata che dice +"/.  
448 \*LUC: eh [<] .  
449 \*PAO: +" festeggio te invece di mia madre perché mia  
450 madre è a Latina@0 # in Italia .  
451 \*PAO: +" e tu sei qui .  
452 \*PAO: +" quindi tu &f [///] prendi la: +//.  
453 \*PAO: però è stato proprio ## adattato o comunque  
454 manipolato fortemen(te) .  
455 \*PAO: un # infatti è una puntata che non si capisce  
456 niente .

---

435 \*PAO: *for instance going back to The Nanny there's a scene # in*  
436 *which uhm there's the [///] the mother's birthday.*  
437 \*PAO: *and: [/] and basically in the Italian version.*  
438 \*PAO: *I mean that episode does not make much sense because*  
439 *it shows # her while she's preparing the decorations uh@0 +"/.*  
440 \*PAO: *+" happy@en birthday@en mama@en.*  
441 \*PAO: *and the aunt comes.*  
442 \*PAO: *and [/] and the Italian version there's a uhm@0 ## voiceover@en.*  
443 \*PAO: *what do you call it.*  
444 \*PAO: *now I can't think of it in Italian the: [>].*  
445 \*LUC: *yes [<].*  
446 \*PAO: *a [/] a voiceover [>] # that [/] <that*



447            *says> [//] of the [/] of the Tata that says +"/.*  
 448 \*LUC: *well [<].*  
 449 \*PAO: *+” I’m celebrating you instead of my mother because my*  
 450            *mother is in Latina @0 # in Italy.*  
 451 \*PAO: *+” and you’re here.*  
 452 \*PAO: *+” so you &f[//] take the: +//.*  
 453 \*PAO: *however it’s been really ## adapted or anyway*  
 454            *heavily edited.*  
 455 \*PAO: *a # actually it’s an episode where you don’t understand*  
 456            *anything.*

[Manchester 2]

Paolo is talking about the episode of an American TV series (*La Tata*, ‘The Nanny’) which is difficult to understand in the Italian version due to some confusing plot adaptations carried out for the Italian audience. He encounters a problem in line 442, where he interrupts his speech by inserting a hesitation marker and a pause before producing the word *voiceover* as a solution to his search. The narration, however, does not resume after the word is uttered, since Paolo produces further talk in an attempt to correct *voiceover*. He first voices his further lexical search in line 443 (*come si chiama* ‘what do you call it’) offering Luca a chance to intervene and help. He then explains that the problem has to do with the language of the item and his inability to retrieve an Italian equivalent in line 444 (*mo’ non mi viene in italiano* lit. ‘now I can’t think of it in Italian’). He eventually solves the problem in line 446 where, after some initial flagging in the form of retracing, he deploys *voce fuori campo* ‘voiceover’. No sooner has the correction been uttered than the narrative resumes exactly where it had been interrupted – although with some further disfluencies.

The extract represents a further case of an other-language insertion – which has been used at the end of a repair sequence – being itself repaired because it is seen as deviating from the current medium by the speaker who produced it. The repair is self-initiated and self-completed, even though the speaker calls for co-operation from his interlocutor (line 443). The only response that is obtained by Luca, however, is a *sì* ‘yes’ in line 448. While Paolo makes his problem obvious to Luca, there is no overt attempt to collaborate to solve it. Indeed Luca’s *sì*

could indicate that in his view the problem does not need to be solved after all, because he does not feel *voiceover* to be deviant in that context. Luca picks up on Paolo's struggle but does not engage in the repair sequence. Rather, he seems to try and sanction the word *voiceover* to interrupt the repair. The two speakers' alignment to the problem at hand, i.e. the occurrence of the word *voiceover*, is therefore not identical. Even after Luca's acknowledgement of the proposed solution to the lexical search problem, Paolo still has to repair his insertion before continuing with his narrative.

Examples (74) and (75) show how the speakers' respective notions of medium may not always be perfectly aligned; indeed there may be differences in orientation with regards to particular items, which are seen as normative (i.e. part of the medium) by one speaker but deviant by the other. In example (74) a request for repair is presented by the speaker's interlocutor but is treated as a request for repetition, rather than a genuine request for repair. In example (75), the speaker initiates a repair sequence and asks for co-operation, but the interlocutor does not engage in the repair, either because he is unable to (in which case, however, an account for this would probably be expected) or because he is simply trying to terminate the repair sequence, which he sees as not necessary. The lack of alignment between speakers and the negotiation work they carry out support the idea that speakers' categories, hereby, the precise boundaries of the medium they use in a given conversation are not a given but result from interaction.

## 7.4 Concluding remarks

This chapter has looked at the occurrence of flagging in contexts of language alternation from a conversation-analytic perspective. Focus has been on instances of language alternation in conjunction with cases of repair, in particular word searches. From the point of view of the external observer, there are three possible solutions to these problems. Speakers may produce a same-language element (examples 66 and 67), an other-language element requiring no further repair (examples 69 and 70), or another-language element requiring further repair (examples

**71, 72 and 75**). The distinction, however, exists only in the eyes of the analyst (Gafaranga, 2000). In terms of speakers' own categories, only two options exist. Either the solution proposed to the problem is seen as suitable (normative), and as such does not require further repair, or it is not suitable (deviant), and therefore requires further repair work on the part of the speaker and/or the interlocutor. Gafaranga (2000) calls these two solutions "other-language repair" and "medium repair" respectively.

While in the former case speakers are still moving within the boundaries of what they see as the medium of the interaction, in the latter case they are deviating from their own norms or schemes of interpretation of the conversation.

In both cases, flagging before the insertion is instrumental in signalling that the speaker is facing a problem – be it the retrieval of a word or a way of showing awareness about the otherness of the insertion about to be produced. In the case of solutions that are felt as deviant, flagging in the form of explicit accounts and justifications after a solution to the lexical search has been deployed reveals to analysts (and participants also) that the solution proposed was not a suitable one, or at least, it was one which required to be accounted for in order to be sanctioned. The presence of flagging and disfluencies in the repair sequences after a first solution has been produced contributes to show that those sequences had not been previously planned by speakers, but rather are triggered by speakers' awareness of the instance of deviance and their own monitoring of the interaction. This gives an image of speakers as highly sensitive to the norms of the interaction: these norms, however, do not correspond exactly to the external category of language, but are defined within the episode. Elements that may appear similar to the analyst may therefore have different status according to speakers.

In most cases both speakers seem to agree on where the boundaries of the medium lie, i.e. they agree on which elements are normative and which are deviant. Examples (**74**) and (**75**), however, show that this need not be the case, and that elements that are perceived as part of the medium by one speaker may be interpreted as deviant by the other. Single items may become the object of a negotiation between speakers, and bring about a shift of the boundaries of

the medium of the conversation. This is the case in example (74), where *three bedroom house* is accepted by Sara after an unheeded request for repair to her interlocutor.

By providing information on speakers' performance, flagging becomes instrumental in reconstructing and interpreting speakers' orientation to their own speech. This in turns allows analysts to discover the medium of the conversation. Gafaranga (2000; 2007) has shown the non-equivalence of language and medium mostly with reference to multilingual contexts. In particular, he has shown situations in which two speakers use a bilingual medium, represented by the (sometimes) frequent alternation of two languages. In these contexts, the appearance of a third language is normally oriented to as an instance of deviance from the bilingual medium, and is either repaired or interpreted as being functional.

One example of functional deviance is the use of reported speech in a third language from the two used in the medium for mimetic purposes, as in the example in (76). Speakers are using French (in italics) and Kinyarwanda (in plain script). The reported speech segment is in Swahili (bold).

**76.** *Civil war has just erupted in Zaire and participants are talking about the consequences this is going to have on Rwandese refugees in that country.*

A: ubu rero ab (.) [C helping him to wine] buretse (.) abazayuruwa bagiye gutangira ngo (.) **fukuza munyarwanda**

B: //avec raison puisque turi imbwa

A: //xxx (laughter) ariko

C: *avec raison* (.) none se none wanzanira ibibazo iwanjye

A, B, C: (laughter)

---

A: now Zairians Zair (.) [C helping him to wine] wait a minute (.) Zairians are going to start saying **kick out Rwandese**

B: //rightly so as we do not deserve any respect

A: //xxx (laughter) but

C: *rightly so* (.) if you brought problems to my door

A, B, C: (laughter)

(Gafaranga and Torras, 2002: 8-9)

Speakers are using French and Kinyarwanda as the medium of their conversation. The authors argue that this can be inferred from the fact that alternation between the two languages is not flagged, or repaired, nor is it oriented to as deviant. The insertion of the Swahili *fuzuka munyarwanda* 'kick out Rwandese' at the end of A's turn is described by Gafaranga as an instance of 'medium suspension'. The fact that the quotation is not repaired, or does not lead to a change of medium in subsequent turns means that speakers interpret it as functional in the context where it occurred, i.e. that of reported speech. According to Gafaranga, the speaker A quotes what he thinks an imaginary Zairian would say, and uses Swahili to frame his quotation in mimetic terms. In the above example French and Swahili constitute the medium, and all elements from either are interpreted as normative. Insertions from other languages, on the other hand, are seen as deviant.

In the present English-Italian data, by contrast, only two languages are used. Defining the medium of the conversation in our data may therefore pose different challenges. In particular, unlike in multilingual contexts, it may not be possible to ascertain whether a solution to a word search problem is not repaired because it is normative or because, while deviant, it is seen as functional in the context in which it occurs, as in example (76). This difficulty notwithstanding, it remains possible to show which elements are definitely perceived as deviant, as shown by the fact that they are made the object of an explicit repair.

Through the analysis of flagging in conjunction with the repair mechanism, we aimed to show that even in a bilingual context it is possible to show that not all instances of language alternation are equal from the speaker's perspective and that not all other-language elements have the same status as either part or not part of the medium. In the Milan data, for instance, it is possible to see that some Italian insertions into English are perceived as part of the medium (or normative), e.g. in example (69) *quattro sezioni*, 'four forms' *media*, 'middle school' *liceo* 'secondary school' and (70) *cucciolo* 'puppy', while others are oriented to as deviant, such as in example (71) *crisi* 'crisis'. This difference can be seen even within the same interactional episode by the same speakers. Examples (70) and (71) come from the

same conversation and the two Italian insertions are uttered within seconds of one another; while they are both deployed as a solution as part of a repair sequence, they are treated as very different by the speakers, as made clear by the different types of flagging and the presence of further repair after the insertion (in 71). This leads us to argue that, while it may be inaccurate to speak of a bilingual English-Italian medium for that interactional episode, it would be equally inappropriate to talk of a monolingual English-only medium, and automatically consider all Italian insertions as instances of deviance. Rather, it seems that while some Italian elements are perceived as normative, the occurrence of other elements is treated as problematic for the organisation of the conversation and triggers further repair. This differs from the examples provided in Gafaranga (2000), in which, due to the conversation being multilingual, the assumption is that either elements from a certain language are always part of the medium or they are always deviant.

Our data illustrate how in two different groups with similar features, speakers use similar strategies to manage their two languages in ordinary conversation. In particular, they appear to use the same methods of repairing specific problems (word searches) that arise in speech. A close analysis of flagging in conversation reveals to analysts where the boundaries of the medium lie, even in a bilingual situation. As we have already claimed, this entails acknowledging that, from a participant perspective, not all instances of language alternation have identical status. While some are treated as part of the medium, others are oriented to as deviant. At a more general level, this informs our conceptualisation of the notion of medium of the conversation as being neither completely bilingual or monolingual. The evidence showing that participants orient to a category that is intermediate between the two is problematic for proponents of a clear-cut distinction between languages (or codes) in contact situations. ¶

## DISCUSSION AND CONCLUSIONS

The occurrence of flagging or flagged CS has been investigated in two small English-Italian corpora of spontaneous conversation between pairs of late bilingual speakers in Manchester and Milan. The data have been analysed separately from both a structural and a more interpretive standpoint. The structural analysis (chapter 6) is essentially quantitative and has looked at the patterns of flagging in the corpus and the relation between flagging and hesitation/repair in monolingual speech. The interpretive analysis (chapter 7) is essentially qualitative in nature, and has adopted CA methods applied to bilingual data.

In the present chapter, the results from the two analyses of flagging are summarised, after a short review of the overall CS patterns found. Some general implications of the role of flagging for existing work on CS and language contact are then discussed. Particular attention is given to the relationship between flagging and hesitation/repair phenomena in monolingual speech and the distinction between CS and borrowing discussed in previous chapters.

Some methodological implications are examined, with particular reference to the criteria used to classify the data, the size of the sample and the type of speakers under study.

Finally, possible avenues for further research originating from the present work.

## 8.1 Code-switching in the data: a summary

Since the study of flagging is inherently dependent on the presence of CS, a brief summary of data concerning the switches in the data is presented before discussing flagging.

Compared with some of the existing data in the literature, the incidence of CS in the Manchester and Milan data sets is relatively rare. The most prevalent occurrence of the phenomenon in both groups is of the insertional type (Muysken, 2000) and represented by one-word insertions. This is consistent with findings from a number of bilingual corpora (see Poplack, 2004; Matras, 2009) in which one-word insertions constitute the “bulk”, or sometimes the only type, of inserted data. While some instances of alternational CS have also been found, the two analyses in this study have focussed on insertions.

The total number of switches is higher in the Manchester data (225) than in the Milan data (133), meaning that English insertions into Italian (Manchester) are more frequent than Italian insertions into English (Milan). In both groups the principal grammatical categories of switches are the same (lone nouns, multiword NPs<sup>37</sup>, adjectives, verbs). Other grammatical categories, such as adverbs or interjections are only represented by very low figures, sometimes in just one group, and have not been analysed separately.

In Milan, lone nouns alone constitute about three quarters of the insertions. In Manchester lone nouns are also the most frequently inserted grammatical category, but multiword NPs and adjectives represent a larger proportion of the data than in Milan, resulting in a more varied array of insertions. The results from the present data support the hypothesis that, as a general trend, the higher the number of switches in a given data set, the lower the proportion of lone nouns (Gardner-Chloros, 2009).

These differences aside, the picture that emerges is one of relative similarity between the two groups in terms of overall patterns of insertion. For instance, when looking at grammatical categories other than lone nouns, the determiners for multiword NPs are almost always provided by the host language; inserted

---

**37** While lone nouns and multiword NPs both fall within the broader category of noun phrases, they have been kept distinct here because they exhibit different behaviour with regards to flagging (see chapter 5).



adjectives in both groups are generally predicative rather than attributive; and all verb forms are non-finite rather than finite.

In their typology of bilingual speech Deuchar *et al.* (2007) argue that insertion is favoured in recent migrant groups and where there is asymmetry in proficiency in the two languages. Both features are apt descriptors of the Manchester and Milan groups, meaning that the results that were found could be at least partly expected from the characteristics of speakers.

Consequently, the similarity in switching patterns between the groups can also be attributed to similarities in the characteristics of speakers, as highlighted by the results of the questionnaires (see chapter 4). Speakers are not members of a community but rather constitute a loose social network. Besides the lack of a community dimension, speakers are characterised by the late acquisition of the L2, mainstream integration in the host country, relatively short period of residence and generally high levels of education. All of these factors may have an impact on the type of CS produced. Of particular importance is the level of education, which could arguably lead to heightened metalinguistic awareness and more markedly prescriptive attitudes vis-à-vis speech forms, such as CS, which are often stigmatised (see Romaine, 1995).

All of these traits are consistent with the observation, both from the data and questionnaires results, that the speakers do not engage in CS regularly. This had the obvious disadvantage of producing a relatively small number of CS tokens, as noted in chapter 6. In turn, however, the infrequency of CS could be directly linked to the high incidence of flagging, replicating results from previous work (Eppler, 1994; see also Auer, 1999).

## 8.2 Results of the analysis of flagging

The quantitative analysis of flagging is essentially theory-independent, but loosely based on previous work carried out by Poplack and associates on a number bilingual spoken corpora. This approach, referred to by its authors as the “comparative method”, has been originally developed in order to distinguish CS from lexical borrowing. Underlying the comparative method is the hypothesis

that a categorical difference exists between the two phenomena, which would involve different mechanisms of language contact (Poplack and Meechan, 1998: 129). As mentioned in section 2.4, this approach necessarily presupposes that the two languages in CS can be unambiguously distinguished.

Insertions have been classified operationally using a dictionary criterion (see chapter 5). All the insertions that have been analysed were therefore provisionally assumed to be switches. In the Manchester data (L1 Italian) for instance, all English-origin words with entries in an English dictionary but not in an Italian one were treated as English words, while those with entries in both were classified as ambiguous, potentially belonging to both languages.

Previous work using the comparative method has focused mostly on morphological and syntactic measures of integration of lone other-language items into the host language. Turpin (1998) and Jones (2005), by contrast, have also looked at flagging, suggesting that it too could be used as a measure of the integration of switches. The hypothesis is that the more flagged an inserted item is, the less likely it is to be integrated, as suggested in Poplack *et al.* (1989: 393): “the presence of various discourse phenomena in the environment of the [inserted] items [...] might indicate poor integration into host-language discourse.” A similar claim has been made in Hlavac (2011), in which a link is drawn between the occurrence of flagging and the lack of morphological integration of contact forms. Looking at flagging differs from the investigation of morphological or syntactic diagnostic tools in that it represents an eminently discourse phenomenon.

Three main findings emerge from the quantitative analysis of the data as a whole:

1. Speakers in the two groups show similar patterns of flagging (with some exceptions);
2. Flagging is related to the grammatical category of the inserted item;
3. Flagging is related to the absolute frequency of the inserted item in the data.

These results will be presented individually (in sections 8.2.1, 8.2.2 and 8.2.3) before their implications are discussed.

The qualitative analysis of the data, on the other hand, adopts the conversation-analytic approach to CS. Flagging has been studied as revealing of the procedures that speakers use to orderly conduct conversation and manage the linguistic resources at their disposal. What the analyst treats as two distinct languages can be seen by speakers as part of the same medium. The qualitative analysis departs from a radically different standpoint, in that it does not postulate that speakers automatically distinguish between two externally definable languages. Rather, it is argued that by restricting analysts' interpretation and focusing on what participants respond to in interaction, it is possible to discover the boundaries of the medium that they are using and how they manage conversations as an orderly activity. The analysis has revealed that speakers do not treat all other-language insertions as having the same status. While some go unnoticed in speech, others are flagged and sometimes made the object of explicit repair. Flagging plays an instrumental role in marking the boundaries of the conversation's medium, thus revealing to speakers – as well as analysts – the membership to the medium of various inserted items.

### **8.2.1 Similarities between groups**

The strategies used to flag switches and their number and position relative to the switch tend to be the same in both groups. Ante-flagging, i.e. flagging preceding the switch, is preferred to post-flagging, i.e. flagging following the switch.

On some occasions a switch is also flagged both before and after being produced. The preference for ante-flagging is consistent with Hlavac's (2006; 2011) results based on Croatian-English data, which also show that ante-flagging is the most frequently adopted strategy, irrespective of the grammatical category of the inserted item.

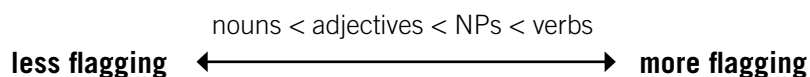
As far as the number of flags is concerned, one single flag (normally preceding the switch) is the preferred type of flagging for both groups and for all grammatical categories with the exclusion of verbs, which always present two or more flags. The types of flags used (unfilled pauses, filled pauses, retracing etc.) also tend to be the same in both groups. One partial exception is the category of syllable

and vowel lengthening, which features more prominently in the Manchester data than in Milan. It should be noted, however, that lengthening of syllables is a conventionalised form of hesitation in Italian, but not so much in English. This cross-linguistic difference alone could account for the difference between the groups with reference to this type of flagging.

In both groups, lone nouns are the most frequently inserted grammatical category. A more in depth-analysis was possible with lone nouns in both data sets, given the higher number of tokens. A comparison was drawn between the rates of discourse phenomena occurring with donor language nouns and a sample of host language nouns (Italian nouns in Italian discourse in Manchester and English nouns in English discourse in Milan). In both groups donor language nouns were produced with significantly higher rates of flagging than their host-language counterparts.

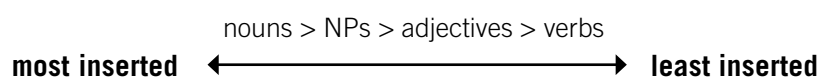
### 8.2.2 Flagging and grammatical category

A relationship exists between the occurrence of flagging and the grammatical category of the donor language item inserted in speech. In terms of percentage, lone nouns are the grammatical category produced with the least flagging. Lone nouns are followed by adjectives, multiword NPs and verbs, which as a category receive the highest amount of flagging (all tokens in both groups are flagged). This pattern is the same in Manchester and Milan. The increasing rate of flagging can be represented as a hierarchy going from the least flagged to the most flagged grammatical category.



Interestingly, if we look at the number of insertions in each grammatical category, regardless of flagging, a very similar pattern emerges. Lone nouns are – perhaps unsurprisingly – the grammatical category inserted most often in both groups. This replicates what has been observed in much previous research on CS data regarding hierarchies of borrowability and “switchability” of lexical items (see

Muysken, 2000 and Matras, 2009 for a survey of these hierarchies in several bilingual data sets), in which nouns consistently appear at the top of the scale. Lone nouns are followed by adjectives, multiword NPs, and verb forms. This pattern is identical in Manchester and Milan. Graphically this would look like the scale below:



Lone nouns are therefore the most frequently inserted and least-flagged grammatical category, while verbs constitute the least-inserted and most flagged category in the data. Multiword NPs and adjectives have intermediate scores between the two, both in terms of flagging rates and raw numbers of tokens.

### 8.2.3 Relation between flagging and frequency

Within single grammatical categories, less frequent items (i.e. with one occurrence in the data) are more flagged than more frequent items (with two or more occurrences). These results show some variation between Manchester and Milan. In order to study this, a frequency boundary was drawn between types with only one occurrence in the corpus (f1) and those with two or more occurrences (f2+). In Manchester, a lower rate of flagging is observed with more frequent (f2+) items compared with less frequent (f1) items. This difference is statistically significant for both English lone nouns and multiword NPs inserted into Italian. If an item belonging to these categories occurs more than once in the corpus, it is less likely to be produced with flagging than an item occurring only once. Adjectives exhibit a similar trend, although they are not amenable to statistical testing due to the small sample size. Verbs are all flagged and are all unique occurrences (f1), meaning that a statistical test was not possible for that category either.

The pattern in Milan is slightly different. Only Italian lone nouns inserted into English show a statistically significant effect of frequency with regard to the overall occurrence of flagging. The same trend (i.e. decrease of flagging with increased frequency) is observable for multiword NPs and adjectives, but is not statistically

significant. As in the Manchester data, all Italian verbs inserted into English in Milan are flagged, implying that further differentiation is not possible. It is worth remembering that the number of overall flagged insertions in Milan is smaller than Manchester to start with, meaning that grammatical categories other than nouns show particularly low numbers, and are not testable even with basic statistical tools. With reference to single nouns alone, this grammatical category was over-represented and thus provided large enough samples in each group for a fine-grained comparison to be carried out. In this category, the data were therefore also classified as low frequency (f1), middle frequency (f2-3), or high frequency (f4+). This analysis revealed that the most frequent insertions (with four or more occurrences in the data) behaved differently from the lower-frequency items. For this group (f4+), there was a dramatic and statistically significant decrease of flagging rates in Milan, while in Manchester the decrease was not quite as sharp. In Manchester even high-frequency single nouns tend to be relatively flagged, while in Milan it appears that after a certain threshold, flagging all but disappears. In general, therefore, increased frequency is related to reduced flagging. This link is stronger in Milan than in Manchester, at least with reference to single nouns (as shown by the difference in rates for f4+ items).

### 8.3 Discussion

The results from the two analyses of flagging have implications for existing theories of CS and bilingualism in general. These are discussed in detail in the present section. The relationship between flagging and hesitation/repair, which had been outlined in chapter 3 is discussed first. The tenability of the CS-borrowing distinction is discussed next, and is followed by some reflections on the differences between grammatical categories in terms of flagging rates. Insights from the CA analysis, which are relevant for the CS-borrowing distinction, are also reviewed.

### 8.3.1 Flagging in code-switching and related discourse phenomena in monolingual speech

In the present work, flagging in CS has been connected to hesitation and repair phenomena in monolingual speech. Our definition of flagging includes several of the devices characterised as hesitation or repair beyond CS. The similarities between the two phenomena are on two levels: at a superficial level, the strategies used in the two contexts are in many cases the same (i.e. unfilled pauses, vowel lengthening, retracing etc.); at a more functional level, the effects that are obtained in monolingual and bilingual speech are also somewhat similar. These include the interruption of speech flow, the correction of a problem, the drawing of attention to the following (or preceding) item inserted in speech or its perceived markedness in context. Both these structural and functional similarities could be taken to argue in favour of the idea that flagging is nothing more than the equivalent of hesitation and repair phenomena in a bilingual environment, i.e. in the immediate proximity of a switch.

This assumption sounds plausible and there is undeniably a certain degree of overlap between the phenomena encountered in monolingual and bilingual speech in our data. By “monolingual speech” here we mean those passages of conversation in which no switches occur. These are treated as an approximation of the monolingual production by the same speakers, following Gardner-Chloros *et al.* (2000). It might be said that some occurrences of discourse phenomena that have been classed as flagging in the present work could have been nothing more than instances of hesitation or repair, and that their occurrence adjacent to a switch is coincidental (i.e. not occurred with a switch *qua* switch).

However, we argue there is definitely an element of specificity in the occurrence of flagging in a bilingual context, i.e. in CS. This is supported by the results of both the quantitative and qualitative analysis of the data. In a purely quantitative sense, the different rates of flagging for donor (switches) and host language (non-switches) lone nouns both in Manchester and Milan support the idea of a categorical distinction between monolingual and bilingual contexts. If the flagging rates had been identical or similar for donor and host language single nouns, then the occurrence of flagging with donor language nouns would have

merely been instances of the hesitation and repair phenomena which are typical of monolingual speech. Since the rates of flagging differ greatly between the two types of nouns, however, there are at least two conclusions that can be drawn: (a) flagging in CS is at least partly distinct from hesitation and repair in monolingual speech; and (b) donor language nouns are categorically different from host language nouns in both groups, in that they show different degrees of integration at the discourse level.

At this point, it would still be possible to interpret this difference as essentially a weaker or stronger manifestation of the same hesitation/repair phenomenon – since the difference is only quantitative. Again with reference to lone nouns, switches are produced with more discourse phenomena than non-switches, but these discourse phenomena tend to be the same as with non-switches.

In order to claim that a difference between the two categories exists, the flagging/hesitation devices produced with switches should be different from those produced with non-switches.

Evidence that the two phenomena are distinct is found in the types of flagging used in donor language nouns (switches), as discussed in section **6.3.1.3**.

In this group we observe a relatively high rate of flagging through metalinguistic commentary (justifications, translation equivalents) which is completely absent with host language nouns. Parallel to this, unfilled pauses, which are the most frequently occurring discourse phenomenon in non-switch contexts, show a sharp decrease. The increased presence of metalinguistic commentary and the fact that the distribution of flags between donor language and host language nouns is significantly different with other flags too corroborates the idea that the occurrence of flagging has a definite element of specificity to CS, and the two phenomena are distinct, at least partly. In Milan the difference in the distribution of flagging types in the two contexts is statistically significant. In Manchester the same difference can be observed between the flags of English (donor language) and Italian (host) nouns, even though the variation in the overall distribution of flag types across the sample is not statistically significant (see section **6.3.1.3**).

It has been argued that hesitation phenomena have a signalling value, even in monolingual speech, because they offer the interlocutor a comment on the



speakers' own performance (see chapter 3). For example they can indicate that the item about to be produced will be delayed, or is infrequent, or may not be appropriate for the specific context. While most of these functions are not self-evident and can at best be inferred implicitly from the context, metalinguistic commentaries provide a very strong clue that the other-language element inserted has indeed a special status. Significantly, with nouns these types of flags are all but absent in non-switch contexts, but appear in both groups in switching contexts. The more explicitly metalinguistic types of flagging become particularly important when the phenomenon is looked at from a CA perspective (discussed in section 8.3.4). The CA analysis has shown the similarity of some forms of the discourse phenomena occurring when speakers repair word search problems occurring in conversation, either with a same language or an other-language element. Additional types of flagging, however, occur when the proposed solution to a word retrieval problem is perceived as deviant by speakers themselves. These more markedly metalinguistic types of flags tend to occur only with some other-language insertions, i.e. in switching contexts. These findings also support the idea that flagging is specific to CS, at least partly, and fulfils additional functions compared with discourse phenomena encountered in monolingual speech.

### 8.3.2 Flagging and the code-switching borrowing distinction

Poplack and associates have argued that, using the comparative method, it is possible to operationally distinguish between CS and borrowing based on the patterning of host language and donor-language lone items in speech, by using a series of morphological and syntactic criteria. As discussed earlier, this presupposes the possibility of unambiguously distinguishing between the two languages. According to this view, if donor language items pattern like host language items in the varieties spoken by the same speakers with regard to the occurrence of specific morpho-syntactic features, they will be considered borrowings rather than code-switches. Criteria used in existing studies include number and gender marking, presence of determiners, adjectival placement and morpho-phonological phenomena. These criteria need to be determined on an *ad hoc* basis for each

language pair and are therefore for the most part dependent on the combination of languages in contact and their specific typological features. For example, if one of the languages has nominal case marking and the other one has not (as in the case of Finnish-English data analysed in Poplack *et al.*, 1989), the presence of case marking with English insertions can be used as a diagnostic tool. In order to be able to contrast the patterns of various groups of items, the two languages should therefore have sufficiently different structures.

In most corpora analysed according with the comparative method, lone nouns in particular have been found to pattern largely like host language items and unlike source language items, thus implying that they are instances of lexical borrowing, rather than CS (see Adalar and Tagliamonte, 1998; Budzhak-Jones, 1998; Eze, 1998; Ghafar Samar and Meechan, 1998). While not a morpho-syntactic feature, flagging has also been used as a criterion by Turpin (1998), who also concludes that the majority of other-language nouns in her French-English data are borrowings and not switches (but see section 3.2.2 for a critique of this interpretation).

In chapter 6, a version of the comparative method was applied to lone nouns, showing that donor language nouns (English in Manchester and Italian in Milan) are significantly more flagged than host language nouns. This is true for both groups. The results from the analysis on lone nouns differ from a number of existing studies, in which donor language lone nouns pattern substantially like host language nouns with reference to the diagnostic criteria adopted. In our data, by contrast, donor language single nouns (switched nouns) exhibit a radically different pattern from host language single nouns (non switched nouns) with reference to flagging. If the similarity of patterning in the two categories is taken as a measure of integration into the host language, this means that there is a large number of single nouns in the present data which are poorly-integrated compared with host language single nouns. The lack of integration does not support the idea that the majority of single nouns are instances of lexical borrowing (whether established or nonce), since these would be expected to behave like host language nouns.

When considered generally, the relation between flagging of single nouns and frequency also argues against the idea that most items are borrowings and that

there is a categorical difference between CS and borrowings in the data. Rather, the fact that flagging gradually decreases with increased frequency of the item implies that there is a continuum of integration at the discourse level. More frequent items are produced more smoothly in speech, and become more similar to host language nouns, or in other words, more integrated. The distribution of flagging therefore supports the idea that CS and borrowing are placed along a continuum (see Myers-Scotton, 1993a; Gardner-Chloros, 2009; Matras, 2009), rather than being two completely distinct phenomena. The separation of languages underpinning the comparative method becomes more difficult to maintain in light of the results from the analysis of flagging, which seems to show a gradient of integration, rather than a clear-cut distinction.

It could be argued that flagging alone is not sufficient to ascertain the integration – or lack thereof – of other-language items into the host language, and that other measures of integration should be used to corroborate or contradict the results from the analysis of flagging. In the present data, however, additional measures would have proven of limited applicability for a number of reasons. Typologically, English and Italian are relatively close. If we look at the morpho-syntactic features of nouns in the two languages, there are many similarities which prevent the application of the comparative method: for instance both languages have both definite and indefinite determiners preceding nouns, and their distribution is partly overlapping, while neither language has overt case marking on nouns<sup>38</sup>. A much larger data set would have been required to find sufficient numbers of lone nouns in contexts where, for instance, one language requires no determiner but the other does, and to see whether single elements pattern with one or the other.

One diagnostic criterion which could have been used to study the integration of nouns is plural marking, since the two languages are sufficiently different in this respect. English normally adds the –s marker (realised as [s], [z] or [ɪz] depending on context) to the singular form, while Italian changes the final vowel if the word ends in an unstressed vowel and leaves the noun invariable in other

---

**38** Arguably English retains genitive inflection with nouns in some cases, but this differs from full case-marking.

cases. So for instance *nome* ‘name’ changes to *nomi* ‘names’, while *città* ‘city’ or *condor* ‘condor’, ending in a stressed vowel and a consonant respectively, remain invariable in the plural. At a superficial level, these differences provide a reliable diagnostic tool, since the two languages have sufficiently different ways of marking the plural. In Manchester we would expect integrated plural English nouns (borrowings) to take a zero plural marking, whereas in Milan integrated Italian nouns would add the English –s suffix to the singular form of the noun. The limitation of this approach is that it would only be possible to test plural nouns – which crucially constitute a small minority of insertions in both Manchester and Milan.

The situation, however, is made much more complex by sociolinguistic and cultural factors. In the Manchester data, we have English insertions into otherwise Italian discourse. Due to the massive influx of Anglo-American loanwords into Italian in recent decades (see section 5.3.2), most monolingual speakers of Italian with little or no functional competence in English will be acquainted with many of these new words, due to their prestige, currency and visibility in the media, advertising, business and at all levels of society (Antonelli, 2007). Italian speakers in Italy sometimes retain English –s plural marking with English-origin words, instead of applying the Italian template, which would lead one to expect nouns not ending in an unstressed vowel to be invariable. The production of English plural morphology is increasingly becoming established as a stylistic marker, by which the speaker signals his/her knowledge of the source-language plural form<sup>39</sup>. Interestingly this has also been recently attested with established borrowings that are prevalently used as invariables, e.g. with dictionary-attested *gli sports*, *i fans* instead of *gli sport* ‘the sports’, *i fan* ‘the fans’ (see Antonelli, 2007). Prescriptive Italian grammarians would expect these nouns to be invariable (see Dardano, 1986), and while this seems to be the preferred option, the situation in contemporary Italian is quite variable, with the two options co-existing even with a number of relatively common loanwords. Variability in

---

<sup>39</sup> Antonelli (2007: 16) also observes how the influence and the prestige attached to English are so strong in contemporary Italian that even loans from languages other than English (including Latin and ancient Greek) are sometimes treated as English and produced with an approximation of English phonology.

plural marking strategies has also been found with English loans in written Italian (Pirkkalainen, 2002), where both options are shown to exist. The fact that even dictionary-attested established borrowings occasionally retain source-language number marking in the speech functionally monolingual speakers of Italian makes number marking a less than reliable diagnostic tool for measuring the integration of inserted English nouns into Italian. This phenomenon could perhaps be seen as an instance of hypercorrection, given the currency and prestige associated with knowledge of English.

For the Milan data, the situation is potentially less ambiguous, since the reverse of the English influence on Italian (i.e. large number of Italian words entering English as loanwords) is not attested. The overwhelming majority of Italian lone nouns in the Milan data, however, are in the singular, and the few tokens of plural nouns all retain Italian (donor language) plural marking. If we were to interpret our results according to Poplack's comparative method, these would provide some evidence that plural Italian single nouns inserted in English are switches rather than borrowings.

### 8.3.3 Differences between grammatical categories

In the quantitative analysis in chapter 6, lone nouns and multiword NPs have been treated as two distinct categories, even though from a syntactic point of view, they are both types of noun phrases. The reason to keep these two groups separate is that they exhibit quite different behaviour with regards to flagging. As mentioned above, a relation holds between the grammatical category of the inserted item and the presence of flagging. Results from the analysis, however, also show that differences can be found between subtypes of the same category. In both Manchester and Milan, lone nouns are produced with 43% flagging, while multiword NPs have 60% flagging in Manchester and 65% in Milan. The difference between lone nouns and multiword NPs is statistically significant in Manchester, but not in Milan. If lone nouns and NPs were inserted in speech in exactly the same way, we would expect them to have similar rates of flagging. The fact that in both groups a considerably higher percentage of multiword

NPs is flagged compared with single nouns, however, suggests, that there is a difference between the two categories.

If the increased presence of flagging is to be seen as a signal, it is possible to argue that speakers encounter more difficulties when inserting multiword NPs compared with lone nouns, or that they perceive them as more obtrusive forms (or less appropriate in the context). Speakers are therefore more likely to forewarn their interlocutor or showing their awareness of the inserted segment than when inserting a lone noun. The two possible explanations (increased difficulty or increased awareness) are in fact interrelated.

The insertion of a multiword NP could entail the activation of structure from the donor language, more so than when inserting a lone noun. The Matrix Language Frame Model (Myers-Scotton, 1993a; 2002) argues precisely that multimorphemic<sup>40</sup> insertions that show structure from the donor language form so-called “Embedded Language Islands”, which entail temporary higher activation of that language. The insertion of these islands may prove more complex for speakers to produce compared with the insertion of lone nouns, because of added grammatical requirements from the donor language – for instance, head-modifier order which is incompatible with the recipient language or different morphological marking system.

The difference in patterns observed between lone nouns and multiword NPs also raises issues for the hypothesis that the overwhelming majority of insertional switches are produced as lexical units or “chunks” (Backus, 2003). Backus’ argument is that most multimorphemic elements are stored as units in the speaker’s lexicon, and are therefore retrieved as a whole rather than morpheme by morpheme.

As already argued in the quantitative analysis in chapter 6 it is undeniable that some multiword NPs are inserted into the host language as units and are not produced synchronically morpheme by morpheme. This is also supported by their semantic interpretation, which is not entirely compositional, i.e. derived from the sum of the components, but is partly idiomatic. This is the case of

---

<sup>40</sup> The label “multimorphemic unit” used by Myers-Scotton and Backus is broader than the label “multiword unit” used here. A lone inserted noun with plural inflection could also be interpreted as a multimorphemic unit (and therefore an Embedded Language island) as being made up by a head noun + a morphological marker (see also Backus, 2003; Deuchar, 2005).

insertions such as *scuola materna* 'kindergarten' in the Milan data or *tuition fees*, *summer school* in Manchester, whose meaning conjures up a unitary concept which cannot be derived exclusively from the juxtaposition of the constituent morphemes. If most insertions were units, it would be difficult to justify how, even within the category of the determiner phrase, a difference in behaviour is observed between lone nouns and more complex lexical units, with increased flagging with the latter. If multiword NPs were units, it would be unreasonable to expect a difference in behaviour between the two groups.

Our results, therefore, are problematic for Backus' claims. While it can easily be conceded that at least some multiword insertions in the present data should be classified as units, (because their semantic interpretation is not strictly compositional) for others a more compositional account should be favoured. The fact that not all insertions are produced equally smoothly introduces a difference within a grammatical category (i.e. determiner phrases) which is not easily reconciled with Backus' hypothesis about the unitary status of the majority of switches.

It should be borne in mind, however, that the occurrence of flagging is not a clear-cut predictor of whether single insertions – be they lone nouns or multiword NPs – possess unit status or not, but should be seen as the indicator of a trend. When looking at the behaviour of grammatical categories in the aggregate, therefore, we argue that the presence of flagging can be indicative of a certain degree of morpho-syntactic complexity and difficulty of insertion. The high degree of flagging observed with verbs in both data sets also supports this idea. Verbs entertain a series of complex morpho-syntactic and semantic relationships in the clause that they appear in (for instance they assign thematic roles and place subcategorisation restrictions in the contexts where they occur) and their insertion may prove to be more difficult for speakers than that of a noun or adjective. Differences between the scant inflectional morphology of English verbs and the complex systems of Italian verbs may also be part of the reason why verbal insertions are always flagged in the data.

### 8.3.4 A note about qualitative analysis

The results from the qualitative analysis in chapter 7 were obtained by analysing the data from a radically different standpoint from the quantitative analysis, which has been the main focus of the discussion so far. While the two analyses are derived from very different theoretical premises, there are nonetheless aspects of the analysis carried out following a CA perspective which can corroborate some of the points made so far about the role of flagging in CS.

One of the main contributions of the CA approach to the study of CS data is that it contributes to overcoming the impasse of whether single instances of language alternation can be considered as instances of lexical borrowing or switches (Pasquandrea, 2007), because the existence of two discrete linguistic systems (such as “English” and “Italian”) is not assumed *a priori*. Rather, the point of departure is that speakers use a medium, which may contain elements from more than one language, as seen by the external analyst. It is by looking at how participants manage specific instances of language alternation that we can try and reconstruct their orientation and the status of specific items.

The qualitative analysis has focused mostly on the mechanism of repair in conjunction with word searches. This is consistent with the interpretation of the discourse phenomena classed as flagging as markers of hesitation or a related problem in speech delivery. In the analysis, flagging is seen as the overt manifestation of the process of repair.

The analysis has shown that speakers can repair the problems arising in speech in a variety of ways. More than the language of the item deployed as a solution to the problem, which is not necessarily seen as meaningful by speakers, it is the presence and the type of flagging which provides essential information to the interlocutor (and the analyst) regarding the status of the item as being part of the medium or an instance of deviance from it. In particular, the presence of metalinguistic commentary (pre-empting, justification, translation equivalents) indicates that the inserted item is not perceived as part of the medium by speakers themselves. This supports the idea that, in some cases at least, flagging is very much a conscious strategy adopted by speakers, as it demonstrates awareness of the insertion of other-language elements perceived as deviant.



Crucially, the fact that not all other-language elements are treated in the same way in discourse invites more caution regarding the possibility of ascribing single items to either one or the other language, given that speakers are shown to not necessarily attend to the category of language. Rather, by showing how, even within the same conversation, different speakers can orient differently to particular inserted elements through the use of flagging is evidence that the boundaries of the medium are constantly achieved and negotiated in interaction, and are not given *a priori*. In turn, this would argue against the possibility of distinguishing categorically between various outcomes of language contact, such as borrowing and CS.

## 8.4 Considerations on the methodology

Having presented the results of this study and their implications, there are some reflections on the methods adopted which should be borne in mind when situating this work within the body of existing research on CS.

In terms of comparability with previous studies on flagging (notably Turpin, 1998 and Jones, 2005), it should be stressed that different data have been extracted and different criteria have been used in the classification of other-language material in the present study. Jones reports using all contact forms in her data irrespective of grammatical category, while Turpin focuses on lone nouns exclusively. This makes a comparison with previous studies more difficult to establish when it comes to the rate of flagging, since – as has been shown – there is variability between different grammatical categories. Both the above studies consider all inserted donor-language elements as falling into the categories of either lone switches or borrowings. As long as an item is recognised as being donor-language by the analyst, therefore, it has been counted either as a potential switch or loanword. Turpin distinguishes the two by using the comparative method, while Jones separates switches from borrowings based on frequency of occurrence in the corpus.

While, based on our results, we would reject the existence of an absolute boundary between CS and borrowing, other criteria for distinguishing between

the two phenomena have been proposed in the literature (see Stammers, 2009 for a comprehensive review, including both linguistic and usage-based criteria). Alongside the use of the comparative method Poplack (2004) indicates the general features of recurrence and diffusion in a community (i.e. frequency in the data and widespread use by a significant number of speakers) as defining characteristics of borrowings; Cheshire and Gardner-Chloros (1998), on the other hand, indicate the preference for an other-language item over a ‘native’ term by a group of speakers. Both criteria have the advantage of tapping directly into speakers’ usage patterns, but require a rather large data set in order to be reliably used.

In the present study, on the other hand, a dictionary criterion has been adopted (see section 5.3.2) to operationally classify the data, following Myers-Scotton (2002) and Deuchar (2006). All items that are attested in both Italian and English dictionaries have been excluded from the analysis, as they have been interpreted as belonging to both languages, i.e. being established borrowings.

In the case of the Manchester data, this means that a large number of English-origin words have been excluded, because they are already considered to be integrated borrowings. A smaller number of Italian-origin words in the English dictionary have similarly been excluded from the Milan data. Only donor-language words not attested in the dictionary of the host language have therefore been counted as switches. This is in contrast with the aforementioned data on flagging, which did not operate a dictionary-based selection (but see Deuchar (2006) and Stammers (2009) for an application of the dictionary criterion to Welsh-English CS data. In Stammers (2009), in particular, attestation of English-origin verbs in Welsh dictionaries was used as a variable when classifying contact forms).

In spite of its shortcomings, the dictionary criterion has the unquestionable benefit of establishing an operational cut-off point when classifying the data. With reference to Turpin (1998) and Jones (2005), it is reasonable to question what is meant, for instance, by “English-origin” or “contact forms” and how the line is drawn in ambiguous cases. In the present data, the idea that all English-origin forms in Manchester data should be included in the count of potential switches would be too permissive, and inevitably include words that are routinely used even by Italian monolinguals with virtually no access to or proficiency in

English. The dictionary criterion offered a solution to this problem, by allowing the exclusion of those words that are potentially already part of monolingual Italian speakers' repertoire.

A related methodological consideration is that of how far analysts should go when considering what constitutes a contact form. If all Italian-origin forms in English in the Milan data were counted, words such as *piano*, *solo*, *torso*, and *finale*, should also have been included as switches<sup>41</sup>. These are all Italian-origin forms and they have undergone relatively little phonological adaptation compared with the original Italian forms. If one were to go even further, also words such as *novel* (it. *novella*), *portfolio*, (it. *portafogli*), *balcony* (it. *balcone*), *bronze* (it. *bronzo*) and *disaster* (it. *disastro*) should be counted as contact forms because they are "Italian-origin", in spite of phonological and morphological integration into English.

Many of the above words are perhaps extreme cases and their status as legitimate English words is beyond dispute. When dealing with more recent contact forms in a synchronic perspective, or with forms that are only specific to a restricted group of speakers, on the other hand, it is considerably harder to determine what should be counted as part of the host language and what should not. The dictionary criterion offers a relatively straightforward solution, in that it applies a clear-cut distinction and excludes the words which, ideally at least, are already quite widespread in the host language.

One final consideration is that the use of the dictionary criterion may have proven less reliable for the Manchester data (L1 Italian) than for the Milan data (L1 English). Dictionaries are inevitably biased towards written rather than spoken language, so it is to be expected that they will offer a more accurate picture of the written than the oral language. The presence of English loanwords in Italian is most prominent in the written medium (Antonelli, 2007), and the presence of Anglicisms in Italian dictionaries reflects this trend. In the spoken language, on the other hand, the penetration of English elements is fairly low, as argued by De Mauro (1993, in Antonelli, 2007). It is therefore possible that several English words that have been excluded from the present analysis were counted as attested in

---

<sup>41</sup> These words are not present in the data, but are being used to illustrate the argument.

Italian because they appear in written Italian, although not necessarily also in the spoken language. For the Milan data, on the other hand, this issue does not seem to be as pressing, since the presence of Italianisms is rather circumscribed.

The adapted application of Poplack's comparative method also raised some issues due to the features of the data. The method involves comparing the patterns of lone inserted items with equivalent items from both the donor and host language. In the case of Milan data this would have entailed comparing the rate of flagging of Italian nouns in English discourse with both English nouns in English discourse and Italian nouns in Italian discourse by the same speakers. The latter context was not sufficiently available in our data since speakers in Milan overwhelmingly used English (their L1) with occasional insertions from Italian. While there are some instances of intersentential switches and full utterances in Italian, there is far from enough material to extract a sample of unmixed Italian speech. *Mutatis mutandis*, the same considerations apply to Manchester data. This raises the question of whether Poplack's approach places too high requirements on the data, making her approach testable only in those cases in which both languages are sufficiently used in the same conversation. This could have possibly been the case in groups with a more balanced proficiency in the two languages, such as the communities analysed in Poplack (1988).

Poplack's argument for making a comparison with items from both languages is consistent with the assumption that it is always possible to draw a clear and categorical distinction between items belonging to one language and those belonging to the other. Aside from the impossibility of making the same comparison, the rationale in the present work, by contrast, was simply to ascertain whether donor language items behave like host language items with reference to flagging. With lone single nouns at least, it has been shown that this is not the case. A further methodological issue that should be discussed has to do with the size of the sample. For some of the grammatical categories under analysis (namely verbs), the size of the sample (five tokens in Manchester and four in Milan) is perhaps responsible for the lack of any clear flagging patterns.

A disclaimer should also be added about the related concept of frequency of inserted items in the data. When discussing more frequent items as being

produced with less flagging, the frequency is to be intended exclusively as frequency in the data. The two corpora have been collected as a “snapshot” of two specific groups of people; there is therefore no claim that the frequency of a given inserted item in the data is an indicator of either a) the frequency or diffusion of the item in the speech of the group of speakers in general; or b) the frequency or diffusion of the item in the source language in general. The size of the data entails that there is a definite idiosyncratic component in the sample: the repetition of an item within the same conversation is sufficient to classify it as a higher-frequency item – this was indeed the case on more than one occasion. Furthermore, some cases of items exhibiting a high frequency in the data could be due to the use of exact repetitions as a means of establishing cohesion in discourse, as mentioned in section **6.3.3.1**.

A final consideration, besides the size of the data, has to do with characteristics of the speakers and the type of data they produce, already touched upon section in **8.1**. As noted in chapter **4**, we are not dealing with traditional bilingual communities, but with a rather loose cluster of speakers with similar language backgrounds. While speakers are generally highly proficient in their respective L2 and should not be considered language learners, it should be remembered that many of them have acquired their second language at least partly through explicit instruction, alongside immersion in the monolingual linguistic majority in their host country. This, together with the high level of education of the speakers is certainly one of the factors accounting for the relative infrequency and type of CS encountered (i.e. unintegrated single word insertions). On the other hand, the low frequency of switches could be related to higher levels of flagging. The quantitative analysis has shown how this is true of single switch types and grammatical categories, but the same could be said of switches as a whole. As a hypothesis, we could argue that speakers who code-switch more frequently could produce less flagging as a whole than the speakers under analysis here.

Our data contrast starkly with the type of language contact common in traditional Italian migrant communities (see section **2.5**) in which the majority of English contact forms present a high degree of morpho-syntactic and phonological integration of into the host language. While no data is available on

the occurrence of flagging for those speakers in order to make a comparison, Hlavac's (2011) study on second-generation Croatian-English bilingual speakers shows how morphological integration is linked to lower levels of flagging, in a community in which switches are frequent and CS is relatively established as a discourse mode.

## 8.5 Suggestions for further research

The present study has offered a critical understanding of the nature of flagging in the speech of two groups of English-Italian bilingual speakers. While several of the questions that we set out with have been answered, there are many more that have arisen from the results of the present work, and that could be investigated in future studies. These include work with new data as well as the study of flagging from different theoretical approaches.

### 8.5.1 A larger data set

An obvious avenue for future research would be to extend the size of the two data sets, in order to resolve to some of the issues encountered in the quantitative analysis with categories other than lone nouns. In particular, a larger corpus would enable us to see more structured patterns with adjectives and verbs, and verify whether a relationship between flagging and frequency can be discovered for these categories also.

In Manchester, it would be particularly interesting to compare the existing data with data from heritage speakers of Italian belonging to the traditional Italian émigré community, in order to see how extra-linguistic factors affect the occurrence and/or type of flagging as well as the general CS patterns, as mentioned in section 8.4. Based on the discussion in the previous section, we would expect these speakers to produce more morphologically integrated forms, less flagging and a more varied array of switches, not limited to one-word insertions. The greatest difficulty with this comparison would probably be speaker recruitment, since the community has lost its geographical concentration

and language transmission rates seem to have inevitably declined after the second generation, even with more close-knit UK communities (see Tosi, 2001). In Milan this comparison would not be possible due to the absence of a historic migrant Anglophone community. Additional data in which the speakers are using mostly their L2 would also enable a more accurate application of Poplack's comparative method. In order to obtain these data, however, informants would probably have to converse with non-members of their own group (i.e. L1 speakers of the other language), given the prevalence of the L1 in in-group communication.

### **8.5.2 Additional types of flagging in naturalistic data**

A direct extension of the present work would involve looking at other phenomena not analysed here which could also function as flagging. In particular, intonational/prosodic cues and gestural/proxemic information, which have been excluded from the present analysis, could be investigated. Insertions of other language elements have sometimes been observed to be accompanied by a variation in pitch or rising intonational prosody (Kinder, 1988). These suprasegmental features, therefore, could also be looked at as an additional type of flagging. Studying suprasegmental features would involve re-encoding the data to include at least some prosodic information (in CA transcripts this is done impressionistically). Alternatively, it would be possible to carry out an acoustic analysis of switches in order to ascertain whether individual speakers exhibit changes in their pitch or prosodic contours in conjunction with insertions – although this may be problematic to achieve since data recorded in naturalistic settings may not always be of a high enough quality. An analysis of this type should also include a more thorough encoding of pause length, which for reasons of convenience was purely impressionistic.

With regard to the analysis of gesture, gaze and facial expressions, video recordings are sometimes used in the study of naturalistic speech, given the wealth of non-verbal paralinguistic information which is encoded through non-verbal means. Some qualitative work using video recording in the study of CS has been carried out in the field of language in the classroom (Greer, 2008) and

in semi-naturalistic settings (Gullberg *et al.*, 2009). More work could be carried out in a naturalistic setting, although the presence of a camera may prove more obtrusive for informants than an audio recorder and have a heavier impact on the spontaneity of the data. There are also additional ethical considerations to be taken into account; the use of video recordings would make it much more difficult to protect speakers' anonymity, which in turn may also make recruitment and data collection more demanding.

As far as these additional non-verbal types of flagging are concerned, intuitions based on anecdotal evidence would predict that prosodic and gestural flags would sometimes occur simultaneously with the verbal flags analysed in the present work, whereas in other instances they could be the only manifestation of flagging.

### 8.5.3 The role of frequency vs. categorical differences

In the discussion, grammatical category and frequency were treated as separate factors with relation to the presence of flagging. In fact, it is possible that the two factors are related and the differences between categories could be explained by referring to frequency alone. Within single grammatical categories (lone nouns and multiword NPs) less frequent items are more flagged than more frequent items. A larger sample may allow to test whether the same is true of adjectives and verbs as well. More importantly, it would also be possible to statistically test whether the rate of flagging is independent of the grammatical category of the items and is explainable through frequency of occurrence alone. Intuitively, this hypothesis is plausible: other-language adjectives and verbs are inserted less frequently and with more flagging than nouns, but in general speakers produce nouns more frequently than they produce adjectives or verbs; their lexicons will also arguably contain more nouns than tokens of the other categories. Having a lower frequency, items from categories other than nouns may be expected to be produced with more flagging. Semantic considerations may also have an effect of this. Nouns have a higher semantic specificity<sup>42</sup> (Backus, 1996; 2003) which means they may be more frequently inserted in CS than adjectives or verbs.

---

<sup>42</sup> Arguably semantic specificity is related to the idea of CS for the *mot juste*.



#### 8.5.4 Flagging as a psycholinguistic phenomenon

A final area of investigation that may prove particularly insightful for the study of flagging is that of psycholinguistic research, and in particular of bilingual speech production models, which have not been dealt with in the present work. Having established that, as a trend, more frequent items are produced with less flagging than less frequent items, further work could explore the role of frequency on the concepts of entrenchment of items in the bilingual lexicon and their accessibility in speech production. Flagging has been related to hesitation phenomena in monolingual speech, which in turn could be investigated with reference to speakers' fluency rates in each of their languages. Some work exists on the rates of discourse phenomena produced by bilinguals speaking each of their languages separately (Fehringer and Fry, 2007) but little is known about how the use of two languages in CS may affect this.

Hesitation could also be connected to increased difficulties in lexical retrieval for specific words. A large body of research supports the idea that even when speaking monolingually, bilinguals activate both their languages to an extent (see Bialystok *et al.*, 2009 for a review). In experimental settings, bilinguals have consistently been shown to be slower than monolinguals in naming tasks, and also that switching from one language to the next has a cost in terms of a delay in production (Costa and Santesteban, 2004). All these findings point to additional retrieval difficulties for bilinguals. The fact that, in our naturalistic data, a large number of switched items is marked by disfluencies, hesitation and pauses could also be related with some of the findings from this work. The cases of repair for word searches analysed in chapter 7 can be seen as a particularly clear example of possible retrieval difficulties faced by speakers.

Recent work on naming tasks has also shown how language context can influence the speed and ease with which bilinguals switch from their L1 and their L2 in experimental settings (Olson, 2011). In particular, when bilinguals are made to switch frequently they are quicker in naming than when they are only made to switch their languages occasionally. The results lend support to Grosjean's (2008) notion of language modes and the existence of a continuum between an ideal monolingual mode (with complete deactivation of one language) and a bilingual

mode (with equal activation of both languages). Olson (2011) argues that the differences in performances in naming tasks could be due to speakers being in a “more” bilingual mode when they switch frequently than when they switch infrequently. Switching would be more effortful in the latter condition than in the former. The sparseness of switches in the present data and the overall high rate of flagging are consistent with these findings and could be further explored as a way to draw a connection between naturalistic and experimental work. Semi-naturalistic tasks could be designed to induce speakers to switch and compare the features of flagging in this environment with those of flagging in naturalistic data.

## 8.6 Conclusions

This study has aimed chiefly to offer a critical appraisal of flagging in CS. Far from being a mere frill of spontaneous speech, flagging is a complex and patterned phenomenon, which can reveal a wealth of information about the mechanisms underlying CS. The appreciation of the role fulfilled by flagging in CS carries implications for many of our conceptualisations about CS and language contact in general. Flagging represents an under-studied phenomenon within the fields of language alternation and language contact, in spite of a large body of work looking at CS in a variety of settings. One of the reasons for this is probably due to the silent assumption that flagging does not differ at all from discourse phenomena in monolingual speech and is simply a feature of spoken language. Flagging strategies have been analysed in two original data sets of English-Italian bilingual conversation. The data document the linguistic practices of recent migrants forming loose social networks in their host countries. The Milan data in particular document language contact involving English as a minority language, a configuration that has not been widely studied in the existing literature. The study adopted a dual focus, by conducting both an analyst-oriented and participant-oriented analysis on the data. The structural analysis has revealed that the occurrence of flagging is related to the frequency and grammatical category of the switch. The patterning of flagging in the present data is problematic for the idea that a categorical difference exists

between CS and borrowing, and favours the existence of a continuum between the two phenomena, with a gradient of integration. Similarly, the behaviour of flagged switches raises issues for the idea that the languages in a bilingual speaker's repertoires are discrete.

The interpretive analysis has reconstructed speakers' orientation to switches, by looking at flagging through a CA approach. Speakers have been shown to treat phenomena that would normally be classified as homogeneous by the analyst as being substantially different. The occurrence and type of flagging have been interpreted as part of the work that speakers carry out to arrive at a shared meaning in interaction.

More specifically for CA work, the analysis also shows how the notion of medium of the conversation, while largely valid, may need to be refined. The medium of the conversation need not be exclusively bilingual or monolingual (cf. Gafaranga, 2000), but can also be an intermediate configuration for which an *a priori* definition is all but impossible. Instances of misalignment between speakers, made relevant through explicit flagging, have shown how the medium is an interactional achievement, and its boundaries cannot always be determined exactly. Speakers are highly sensitive actors, and a close analysis of how they deal with specific conversational problems through flagging can be particularly revealing for a participant-oriented account of language contact.

The interpretation of flagging as a meaningful interactional feature in CS is not to deny that flagging may well be a signal of disfluency or occasional difficulty in speech production, at least in some cases. What we have aimed to show in the present work, however, is that the acknowledgement of the nature of flagging as a signal of a difficulty does not entail rejecting the idea that the same phenomenon may fulfil other functions in conversation. Some of these functions may be more obvious to speakers – albeit unconsciously – than to the analyst. Yet, this does not mean that we should relinquish attempts to reach a deeper understanding of what routinely goes unnoticed in speech. Adopting an eclectic approach, as in the present study, may be particularly helpful when investigating phenomena which are normally treated as being peripheral. ¶



## Appendix 1

Participant contact letter (bilingual version)

This letter was emailed to potential participants to explain the project. When contact had not been made with one of the participants before the recording, a hard copy of the letter was normally given to them.



PRIFYSGOL  
**BANGOR**  
UNIVERSITY

### **Research Project on Bilingual Communication**

I am a PhD student at Bangor University in Britain and I am currently conducting a study about communication between bilinguals based in Italy and the UK. I am working under the supervision of Professor Margaret Deuchar, the director of the *Centre for Research on Bilingualism in Theory and Practice* at Bangor.

I am looking for people who, like you, regularly speak both English and Italian. For my research project I would need to record an informal conversation between you and another person of your choice – a relative, friend or colleague – who also uses the two languages on a day-to-day basis. Together we can arrange a time and place to meet and make the recording, wherever it is convenient for you (i.e. a café, workplace, etc.) The session will only take up to an hour.

The conversation would just be an informal chat between you and the person coming with you. The topics are up to you. I could however provide you with some ideas just to get started. The recording is not an interview. After the recording session, you will have a chance to listen to the tape in case you are not happy with it and/or you would like sections of it to be edited. I would also ask you to fill in a short anonymous questionnaire.

Once I have your authorisation, I will proceed to analyse the conversation for research purposes. The data will be made anonymous, with names and references to people omitted.

If you have any further questions, comments or complaints about the project, you can contact me by email: [a.rosignoli@bangor.ac.uk](mailto:a.rosignoli@bangor.ac.uk).

Thank you for your co-operation,  
Alberto Rosignoli



PRIFYSGOL  
**BANGOR**  
UNIVERSITY

### **Progetto di ricerca sulla comunicazione bilingue**

Sono un dottorando dell'Università di Bangor, in Gran Bretagna. Al momento sto lavorando a uno studio sulla comunicazione tra persone bilingui, sia in Italia che nel Regno Unito. Il progetto si svolge sotto la guida della Prof. Margaret Deuchar, direttrice del *Centre for Research on Bilingualism in Theory and Practice* a Bangor.

Sono alla ricerca di persone che, come Lei, parlino abitualmente italiano e inglese, per poter registrare una conversazione informale tra Lei e un'altra persona – un familiare, un amico o un collega – che usi quotidianamente le due lingue. Possiamo decidere insieme un orario e un luogo di ritrovo per la registrazione, a seconda delle Vostre esigenze (ad esempio un bar, il luogo di lavoro, etc.). La sessione durerà al massimo un'ora.

La conversazione dovrebbe essere una semplice chiacchierata tra Lei e la persona che L'accompagnerà. Non si tratta di un'intervista: gli argomenti della

conversazione sono liberi; al massimo potrei suggerirvi qualche spunto per cominciare. Dopo la registrazione, sarà possibile riascoltare il dialogo nel caso vogliate modificare o rimuovere alcune sezioni. Come ultima cortesia, Vi chiederò di compilare un breve questionario anonimo.

Una volta ottenuta la Vostra autorizzazione, procederò ad analizzare la conversazione esclusivamente a scopo di ricerca. I dati saranno resi anonimi, omettendo i riferimenti a nomi e persone citate.

Per ulteriori chiarimenti, osservazioni o critiche, non esiti a contattarmi via email all'indirizzo: [a.rosignoli@bangor.ac.uk](mailto:a.rosignoli@bangor.ac.uk).

Grazie per la collaborazione,  
Alberto Rosignoli

## Appendix 2

### Speakers' social networks

The two tables represent the configuration of the speakers' social networks as derived from the researcher's interaction with participants and speakers' recruitment. Bold lines show speakers who have been recorded together.

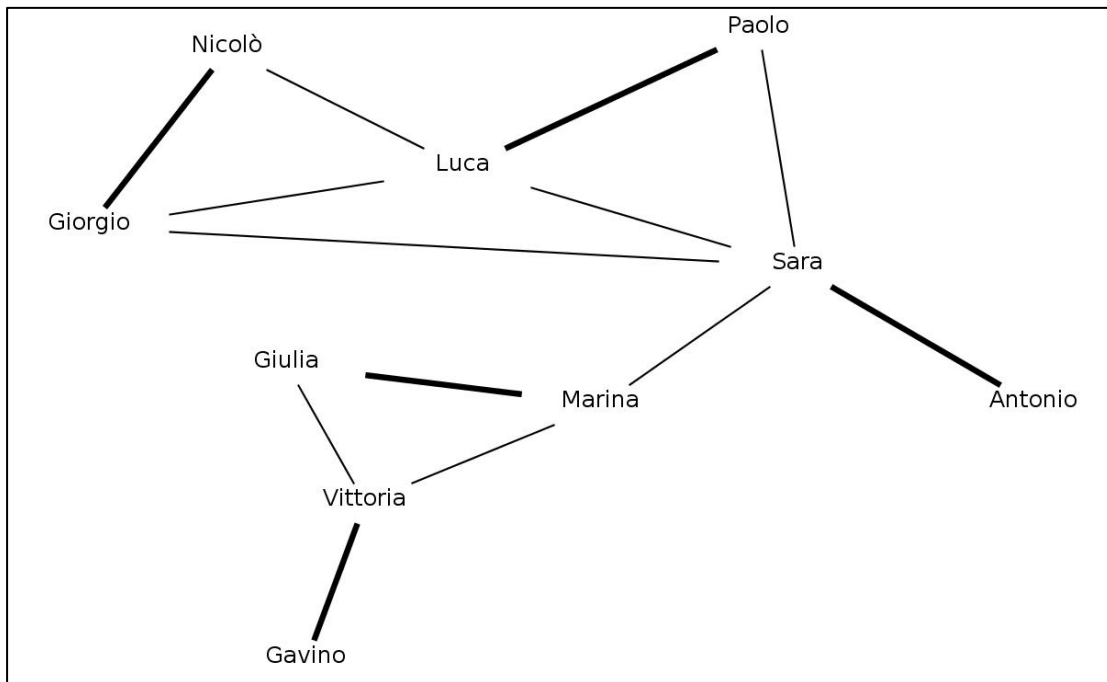


Figure A. Manchester speakers

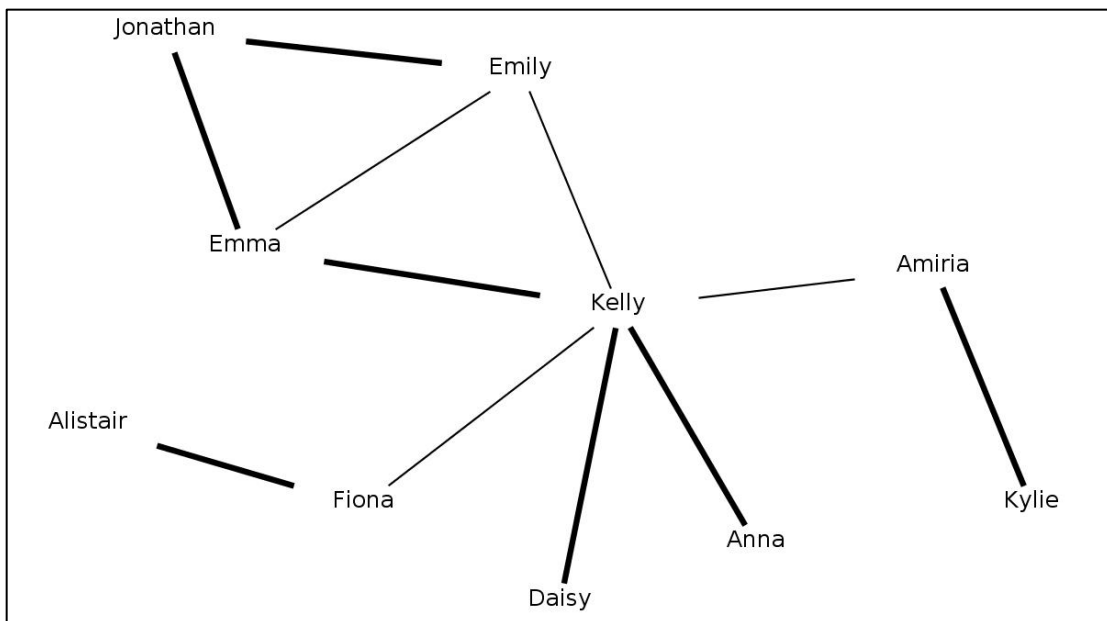


Figure B. Milan speakers



## Appendix 3

### Summary of recordings

Recording	Speaker 1 (alias)	ID1	Speaker 2 (alias)	ID2	date	setting	duration hh:mm:ss
Manchester 1	Sara	MancE	Antonio	MancF	15/11/2088	home	00:54:04
Manchester 2	Luca	MancA	Paolo	MancB	14/11/2008	café	00:54:21
Manchester 3	Marina	MancC	Giulia*	MancD	14/11/2008	café	00:46:34
Manchester 4	Giorgio	MancG	Nicolò	ManchH	06/12/2008	café	00:51:35
Manchester 5	Vittoria	MancI	Gavino	ManchJ	20/02/2009	home	01:04:39
Milano 2	Emma	MilaC	Kelly	MilaF	17/04/2009	work	00:49:08
Milano 3	Amiria	MilaE	Kylie	MilaD	10/06/2009	home	00:37:33
Milano 4	Kelly	MilaF	Anna*	MilaG	10/06/2009	home	00:52:51
Milano 5	Emily	MilaB	Jonathan	MilaA	10/07/2008	work	00:19:41
Milano 6	Emma	MilaC	Jonathan	MilaA	10/07/2008	work	00:35:33
Milano 7	Alistair	MilaI	Fiona	MilaJ	09/06/2009	home	00:46:34
Milano 8	Kelly	MilaF	Daisy	MilaH	11/06/2009	home	00:54:25

\* Simultaneous bilingual.

## Appendix 4

Participant questionnaire (bilingual version)

### Questionnaire

Thank you for participating in the project. I would appreciate it if you could spare a moment and answer these few questions, which will help me with my research. As with all data you provide, please note that your answers will be treated completely anonymously and for academic purposes only.

1. Sex:  Male  Female

2. Year of birth: \_\_\_\_\_

3. Nationality: \_\_\_\_\_

4. Occupation: \_\_\_\_\_

(If you are currently retired or unemployed, please provide what was your last occupation)

5. Please indicate the areas where you have lived for significant periods of your life:

e.g. Place: Milan (Italy) Dates: 1972 - 1985

Place: Manchester (UK) Dates: 1985 - 2000

Place: Rome (Italy) Dates: 2000 - 2008

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

Place: \_\_\_\_\_ Dates: \_\_\_\_\_

**6.** What is the highest level of formal education you have completed?

- GCSE/Junior High/Scuola dell'obbligo or equivalent
- A-Levels/High School/Maturità or equivalent
- Bachelor's/Diploma/Laurea or equivalent
- Master's/Doctorate or equivalent
- None of the above

**7.** Since when have you been able to speak English?

- Since I was 2 years or younger
- Since I was 4 years or younger
- Since primary school
- Since secondary school
- I learnt English as an adult.

**8.** Since when have you been able to speak Italian?

- Since I was 2 years or younger
- Since I was 4 years or younger
- Since primary school
- Since secondary school
- I learnt Italian as an adult

**9.** How well do you feel you can speak English on a scale from 1 to 4?

- 1 Only know some words/expressions
- 2 Confident in basic conversation
- 3 Fairly confident in extended conversation
- 4 Confident in extended conversation

**10.** How well do you feel you can speak Italian on a scale from 1 to 4?

- 1 Only know some words/expressions
- 2 Confident in basic conversation
- 3 Fairly confident in extended conversation
- 4 Confident in extended conversation

11. Can you speak other languages well enough to conduct an informal conversation? Please indicate them here.

---

12. Since when have you been able to speak a language other than English or Italian? (If applicable: please only provide an answer for the language you know best and indicate it here \_\_\_\_\_)

- Since I was 2 years or younger
- Since I was 4 years or younger
- Since primary school
- Since secondary school
- I learnt it as an adult

13. Which language(s) did your mother speak to you as you were growing up (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

14. Which language(s) did your father speak to you as you were growing up (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

**15.** Which language(s) did any other guardian or caregiver speak to you as you were growing up (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

**16.** Through which language were you primarily taught in primary school (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

**17.** Through which language were you primarily taught in secondary school (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

**18.** Through which language were you primarily taught at any other further education institution (if applicable)?

- English
- Italian
- English and Italian
- Other (please specify) \_\_\_\_\_
- N/A

19. In the table below, write the names of the five people you speak most often to everyday, either in person or on the phone (e.g. relative, partner, friend, colleague etc.) then tick the box for the language(s) you mostly speak with each person.

Name of person, or relationship (use fictitious names if you prefer)	Language mostly spoken with that person: (place a tick in one cell below for each line)			
	English	Italian	English and Italian	Other language(s)
1				
2				
3				
4				
5				

21. How would you rate the English language on a scale from 1 to 5 regarding the following properties? Circle one number in each line.

old-fashioned	1	2	3	4	5	modern
unfriendly	1	2	3	4	5	friendly
uninfluential	1	2	3	4	5	influential
uninspiring	1	2	3	4	5	inspiring
useless	1	2	3	4	5	useful
ugly	1	2	3	4	5	beautiful

22. How would you rate the Italian language on a scale from 1 to 5 regarding the following properties? Circle one number in each line.

old-fashioned	1	2	3	4	5	modern
unfriendly	1	2	3	4	5	friendly
uninfluential	1	2	3	4	5	influential
uninspiring	1	2	3	4	5	inspiring
useless	1	2	3	4	5	useful
ugly	1	2	3	4	5	beautiful

**23.** (If applicable) How would you rate one of the other languages you speak on a scale 1 to 5 regarding the following properties? Circle one number in each line (please refer to the language you indicated in question 12).

old-fashioned	1	2	3	4	5	modern
unfriendly	1	2	3	4	5	friendly
uninfluential	1	2	3	4	5	influential
uninspiring	1	2	3	4	5	inspiring
useless	1	2	3	4	5	useful
ugly	1	2	3	4	5	beautiful

**24.** To what extent do you agree with the following statements? Circle one number for each statement.

a. In everyday conversation I keep the Italian and English language separate.

strongly disagree    1    2    3    4    5    strongly agree

b. People should avoid mixing Italian and English in the same conversation.

strongly disagree    1    2    3    4    5    strongly agree

## Questionario

Grazie per la Sua partecipazione al progetto. Le sarei grato se volesse rispondere a queste domande, che mi saranno d'aiuto nel corso della ricerca. Come tutte le informazioni fornite, le Sue risposte saranno elaborate in modo completamente anonimo e verranno utilizzate esclusivamente in ambito accademico a scopo di ricerca.

1. Sesso:                     M                     F

2. Anno di nascita: \_\_\_\_\_

3. Nazionalità: \_\_\_\_\_

4. Professione: \_\_\_\_\_

(se attualmente disoccupato o in pensione, indichi l'ultima professione svolta)

5. Indichi qui di seguito i luoghi dove ha risieduto per lunghi periodi

(escludendo i periodi di vacanza):

Esempio:      Luogo: Roma, Italia                    Periodo: 1972 -1985

                  Luogo: Chester, UK                    Periodo: 1985 -2003

                  Luogo: Bologna, Italia                    Periodo: 2003 -2007

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_

Luogo: \_\_\_\_\_                    Periodo: \_\_\_\_\_



**6.** Qual è il più alto titolo di studio da Lei conseguito?

- GCSE/Junior High/Scuola dell'obbligo o equivalente
- A-Levels/High School/Maturità o equivalente
- Bachelor's/Diploma/Laurea o equivalente
- Master's/Dottorato o equivalente
- Nessuno dei precedenti

**7.** Da quando è in grado di parlare inglese?

- Da quando avevo meno di 2 anni.
- Da quando avevo meno di 4 anni.
- A partire dalla scuola elementare/primaria.
- A partire dalla scuola media/secondaria.
- Ho imparato l'inglese da adulto.

**8.** Da quando è in grado di parlare italiano?

- Da quando avevo meno di 2 anni.
- Da quando avevo meno di 4 anni.
- A partire dalla scuola elementare/primaria.
- A partire dalla scuola media/secondaria.
- Ho imparato l'italiano da adulto.

**9.** In una scala da 1 a 4, quale voto darebbe alla Sua capacità di parlare inglese?

- 1 Conosco solo alcune espressioni/frasi.
- 2 Sicuro in una conversazione semplice.
- 3 Abbastanza sicuro in una conversazione prolungata.
- 4 Sicuro in una conversazione prolungata.

**10.** In una scala da 1 a 4, quale voto darebbe alla Sua capacità di parlare italiano?

- 1 Conosco solo alcune espressioni/frasi.
- 2 Sicuro in una conversazione semplice.
- 3 Abbastanza sicuro in una conversazione prolungata.
- 4 Sicuro in una conversazione prolungata.

11. (Se pertinente) Conosce altre lingue abbastanza bene da utilizzarle per una conversazione informale? Se sì, le indichi qui di seguito:

---

12. (Se pertinente) Da quando è in grado di parlare una lingua diversa dall'inglese e dall'italiano (fornisca solo una risposta, per la lingua che conosce meglio e la indichi qui di seguito \_\_\_\_\_)

- Da quando avevo meno di 2 anni.
- Da quando avevo meno di 4 anni.
- A partire dalla scuola elementare/primaria.
- A partire dalla scuola media/secondaria.
- Ho imparato un'altra lingua da adulto.

13. (Se pertinente) In quale lingua Le parlava Sua madre quand'era piccolo/a?:

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

14. (Se pertinente) In quale lingua Le parlava Suo padre quand'era piccolo/a?

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

**15.** (Se pertinente) In quale lingua Le parlava il Suo tutore/guardiano o altra figura quand'era piccolo/a?

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

**16.** (Se pertinente) Quando frequentava le scuole elementari/primarie in quale lingua si svolgevano le lezioni?

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

**17.** (Se pertinente) Quando frequentava le scuole superiori/secondarie in quale lingua si svolgevano le lezioni?

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

**18.** (Se pertinente) All'università in quale lingua si svolgevano le lezioni?

- Inglese
- Italiano
- Inglese e italiano
- Altro (specificare) \_\_\_\_\_
- N/A

19. Nella tabella qui di seguito, indichi le cinque persone con cui parla più di frequente (di persona o al telefono), dopo di che, contrassegni la casella corrispondente alla lingua (o lingue) che usa normalmente con ogni persona.

Nome della persona o della relazione (se preferisce usi degli pseudonimi)	Lingua parlata più di frequente con la persona: (contrassegni una casella per ogni persona)			
	Inglese	Italiano	Inglese e Italiano	Altro (specificare)
1				
2				
3				
4				
5				

20. In una scala da 1 a 5, quale punteggio assegnerebbe alla lingua inglese riguardo alle seguenti caratteristiche? Indichi un numero in ogni riga.

antiquata	1	2	3	4	5	moderna
ostile	1	2	3	4	5	amichevole
poco autorevole	1	2	3	4	5	molto autorevole
poco suggestiva	1	2	3	4	5	molto suggestiva
inutile	1	2	3	4	5	utile
brutta	1	2	3	4	5	bella

21. In una scala da 1 a 5, quale punteggio assegnerebbe alla lingua italiana riguardo alle seguenti caratteristiche? Indichi un numero in ogni riga.

antiquata	1	2	3	4	5	moderna
ostile	1	2	3	4	5	amichevole
poco autorevole	1	2	3	4	5	molto autorevole
poco suggestiva	1	2	3	4	5	molto suggestiva
inutile	1	2	3	4	5	utile
brutta	1	2	3	4	5	bella

**22.** (Se parla un'altra lingua) In una scala da 1 a 5, quale punteggio assegnerebbe a un'altra delle lingue che Lei conosce riguardo alle seguenti caratteristiche? Indichi un numero in ogni riga (faccia riferimento alla lingua indicata nella domanda 12).

antiquata	1	2	3	4	5	moderna
ostile	1	2	3	4	5	amichevole
poco autorevole	1	2	3	4	5	molto autorevole
poco suggestiva	1	2	3	4	5	molto suggestiva
inutile	1	2	3	4	5	utile
brutta	1	2	3	4	5	bella

**23.** Quanto è d'accordo con le seguenti affermazioni? Indichi un numero per ogni affermazione.

a. Normalmente, quando parlo tengo l'inglese e l'italiano separati.

molto in disaccordo 1      2      3      4      5      molto d'accordo

b. Bisognerebbe evitare di mischiare inglese e italiano nella medesima conversazione.

molto in disaccordo 1      2      3      4      5      molto d'accordo

## Appendix 5

Individual speakers' questionnaire data

Alias	ID	Recording(s)	Sex	Age	Area Brought Up	Edu	English since	Italian since	English Ability	Italian Ability	Mother spoke	Father spoke	Primary School Medium	Secondary School Medium	Higher ed medium
Luca	MancA	Manchester2	M	35	Italy	5	4	1	4	4	IT	IT	IT	E&I	E&I
Paolo	MancB	Manchester2	M	33	Italy	5	4	1	4	4	IT	IT	IT	IT	E&I
Marina	MancC	Manchester3	F	30	Italy	4	4	1	3	4	IT	IT	IT	IT	IT
Giulia	MancD	Manchester3	F	30	Italy	4	1	1	4	4	EN	IT	IT	IT	E&I
Sara	MancE	Manchester1	F	35	Italy	5	5	1	4	4	Paduan	Paduan	IT	IT	IT
Antonio	MancF	Manchester1	M	41	Italy	4	5	1	3	4	IT	IT	IT	IT	IT
Giorgio	MancG	Manchester4	M	33	Italy	5	5	1	4	4	IT	IT	IT	IT	IT
Nicolò	MancH	Manchester4	M	32	Italy	5	4	1	3	4	IT	IT	IT	E&I	E&I
Vittoria	MancI	Manchester5	F	34	Italy	5	4	1	3	4	IT	IT	IT	IT	IT
Gavino	MancJ	Manchester5	M	35	Italy	5	4	1	4	4	IT	IT	IT	IT	IT

Alias	ID	Recording(s)	Sex	Age	Area Brought Up	Edu Level	English since	Italian since	English Ability	Italian Ability	Mother spoke	Father spoke	Primary School Medium	Secondary School Medium	Higher ed medium
Jonathan	MilaA	Milano5/6	M	26	USA	5	1	5	4	3	EN	EN	EN	EN	EN
Emily	MilaB	Milano5	F	38	UK	3	1	5	4	4	EN	EN	EN	EN	EN
Emma	MilaC	Milano2/6	F	36	UK	5	1	5	4	3	EN	EN	EN	EN	E&I
Kylie	MilaD	Milano3	F	41	UK	4	1	5	4	2	EN	EN	EN	EN	EN
Amiria	MilaE	Milano3	F	35	NZ	5	1	5	4	4	EN	EN	EN	EN	EN
Kelly	MilaF	Milano2/4/8	F	24	USA	4	1	5	4	4	EN	EN	EN	EN	EN
Anna	MilaG	Milano4	F	23	Italy	4	1	3	4	4	EN	E&I	IT	IT	
Daisy	MilaH	Milano8	F	26	Italy	3	4	1	4	4	IT	IT	EN	EN	EN
Alistair	MilaI	Milano7	M	37	UK	4	1	5	4	3	EN	EN	EN	EN	EN
Fiona	MilaJ	Milano7	F	40	USA	5	1	5	4	4	EN	EN	EN	EN	EN

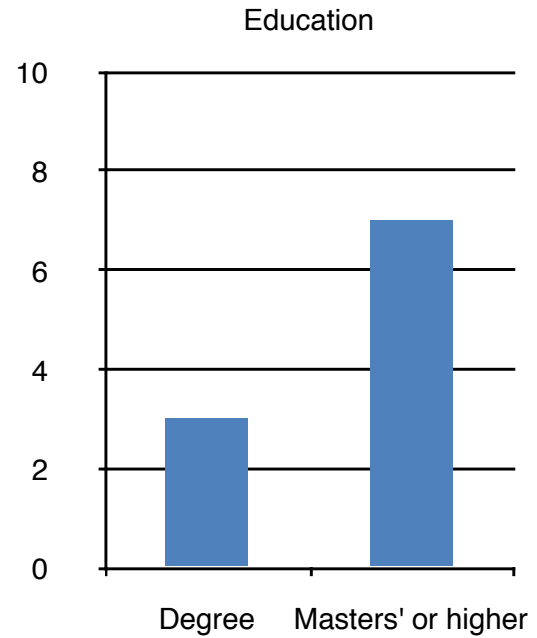
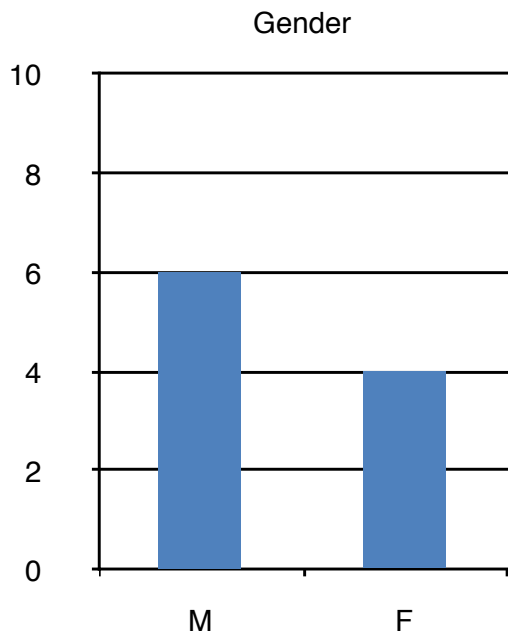
**Education:** **2** GCSE/Junior High/*Scuola dell'obbligo*; **3** A-Levels/High School/*Maturità*; **4** Bachelor's/*Diploma/Laurea*; **5** Master's/Doctorate.

**Eng/Ita since:** **1** two or younger; **2** four or younger; **3** since primary school; **4** since secondary school; **5** as an adult.

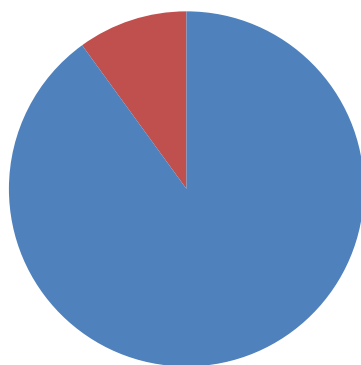
**Language ability:** **1** words/phrases; **2** basic conversation; **3** fairly confident; **4** confident.

## Appendix 6

Manchester speakers' questionnaire data

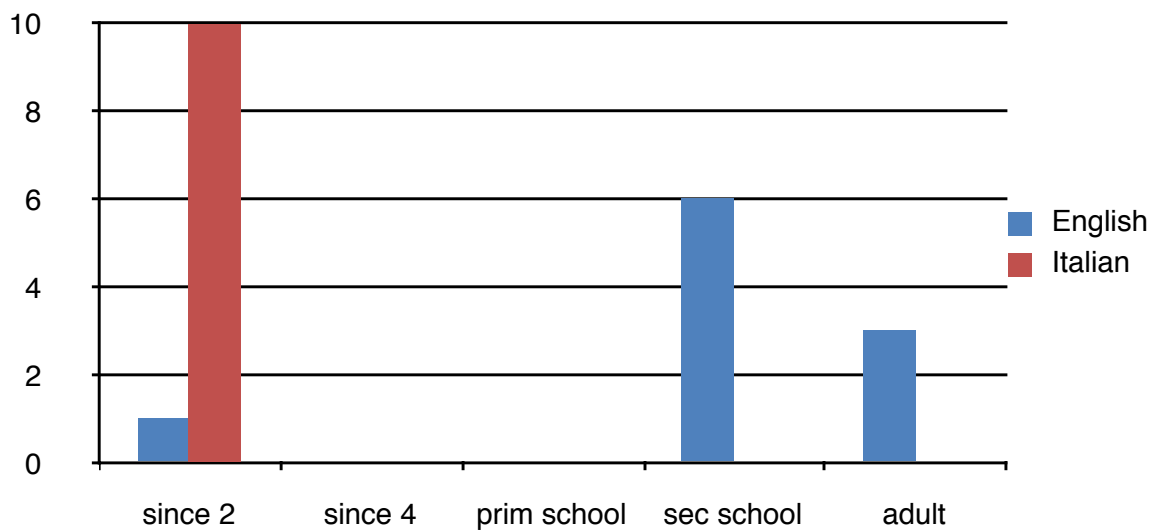


National identity

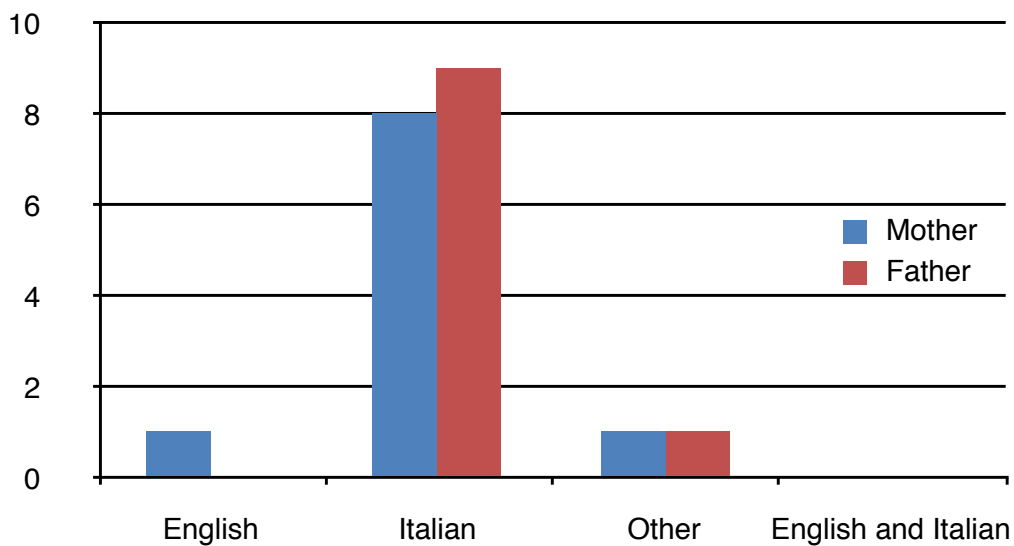


- Italian
- Italian/British

Age of acquisition of English and Italian

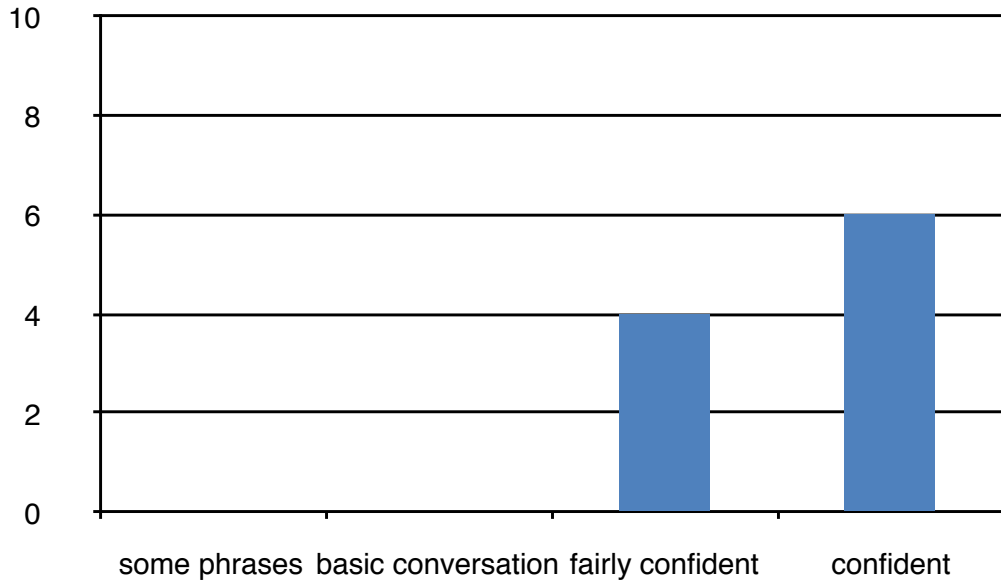


Parental language input

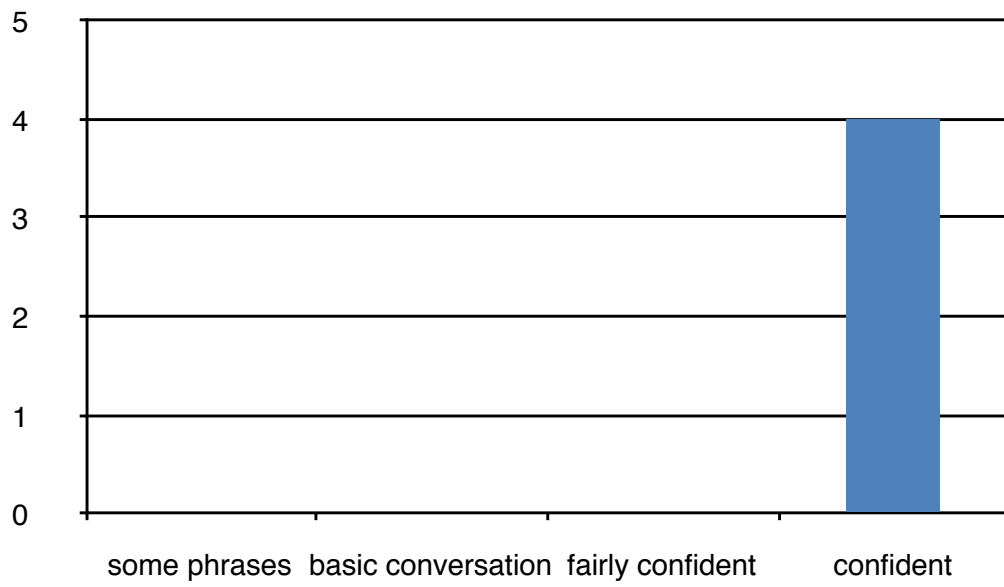


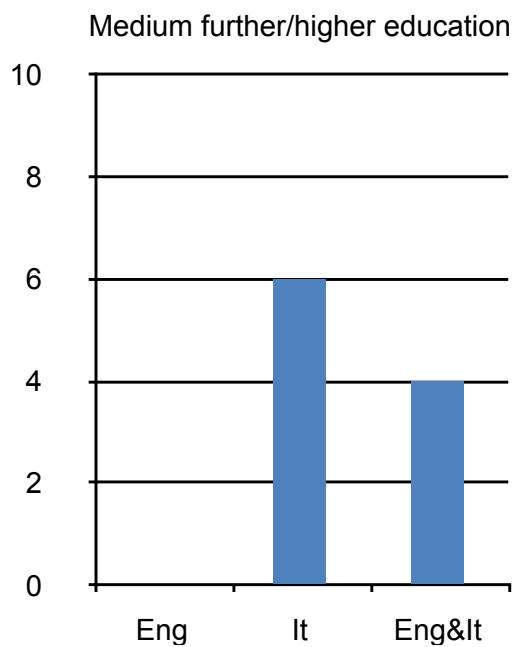
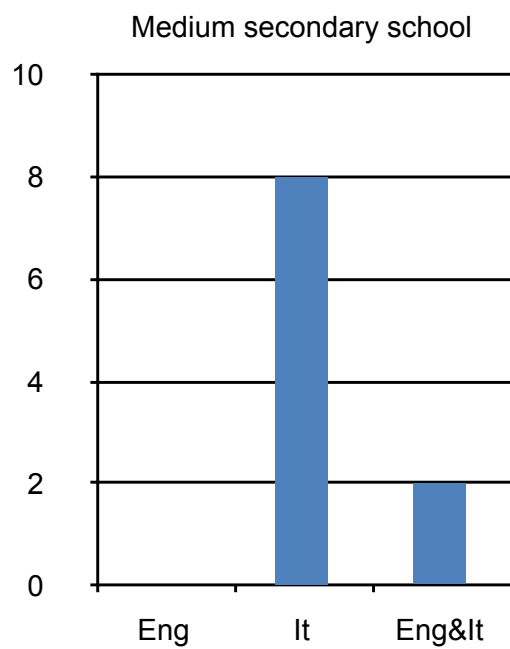
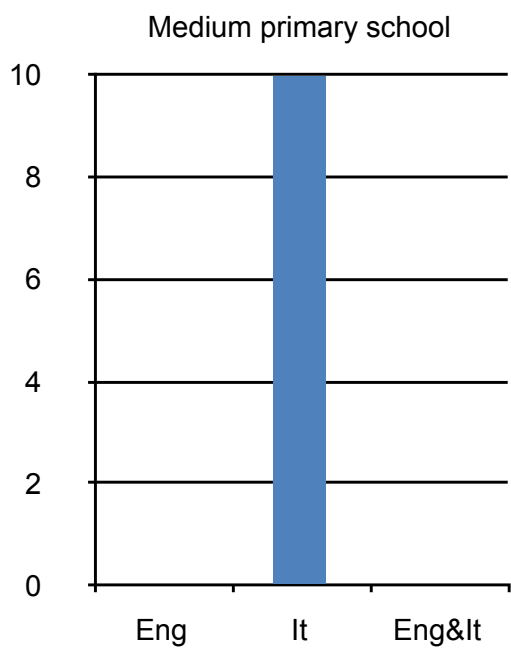


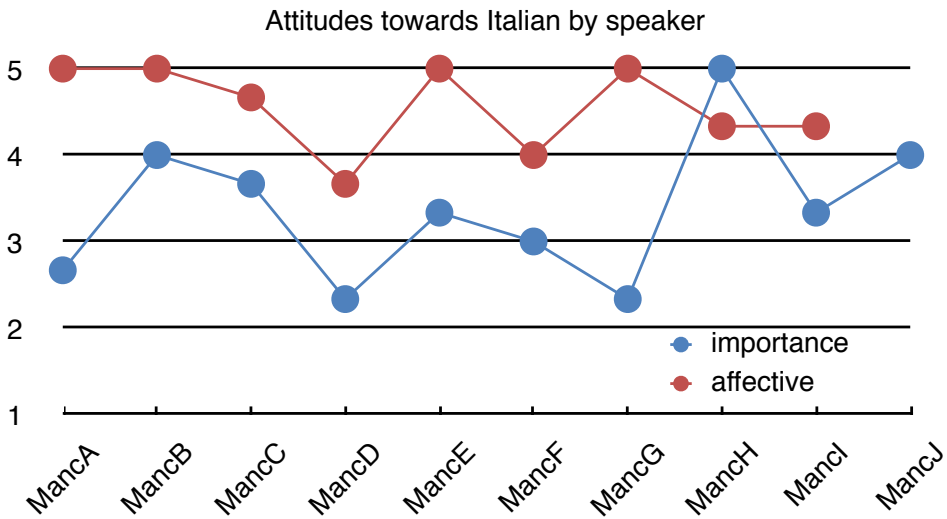
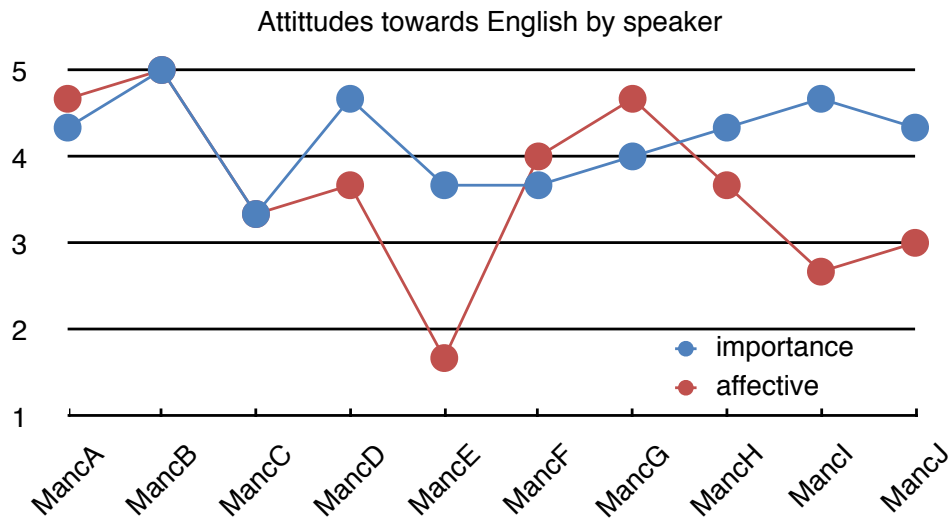
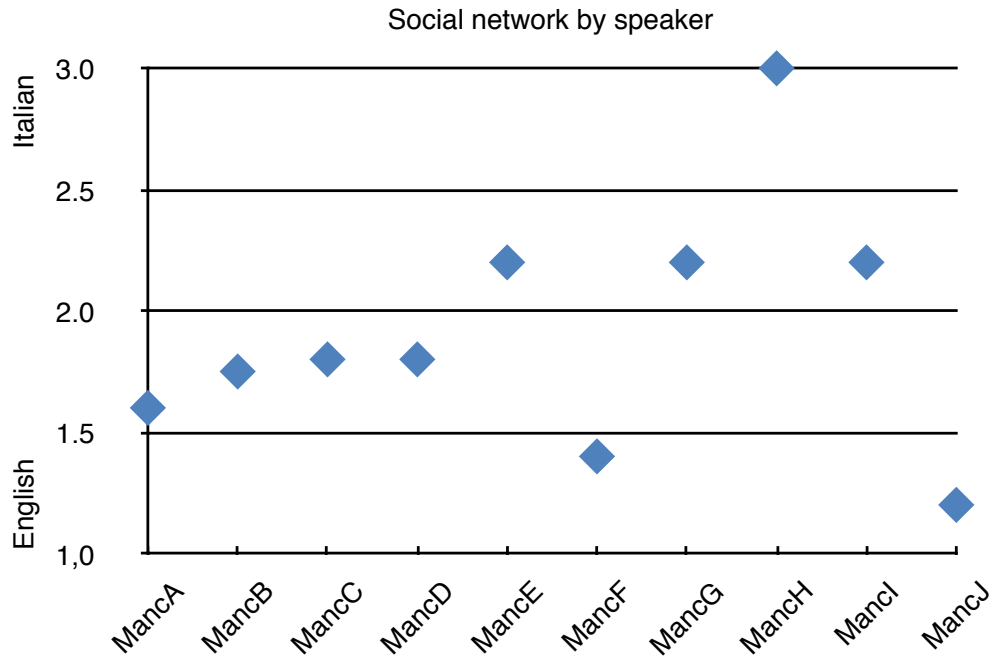
English proficiency



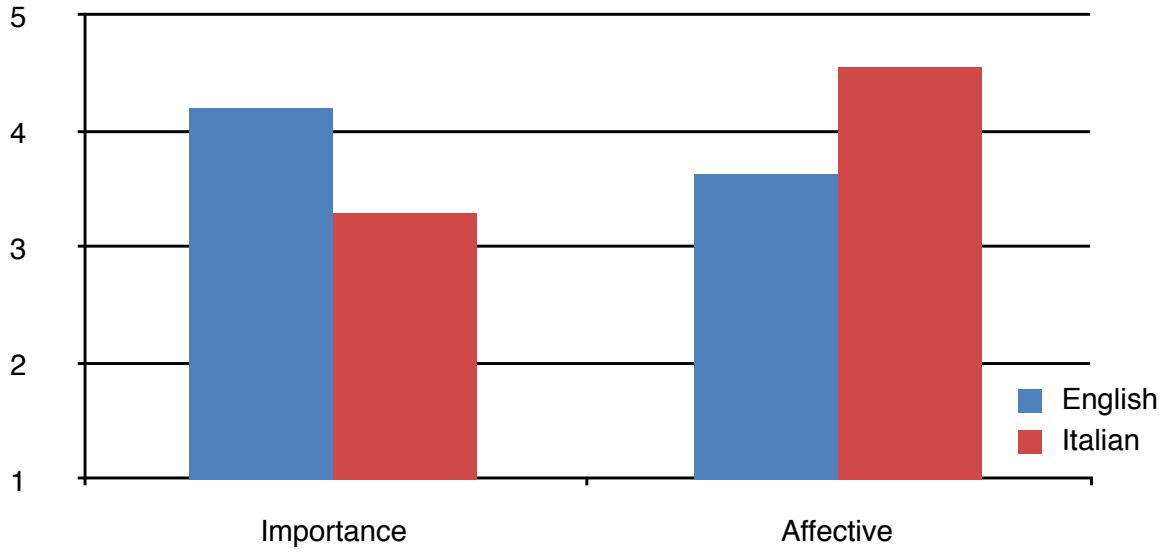
Italian proficiency



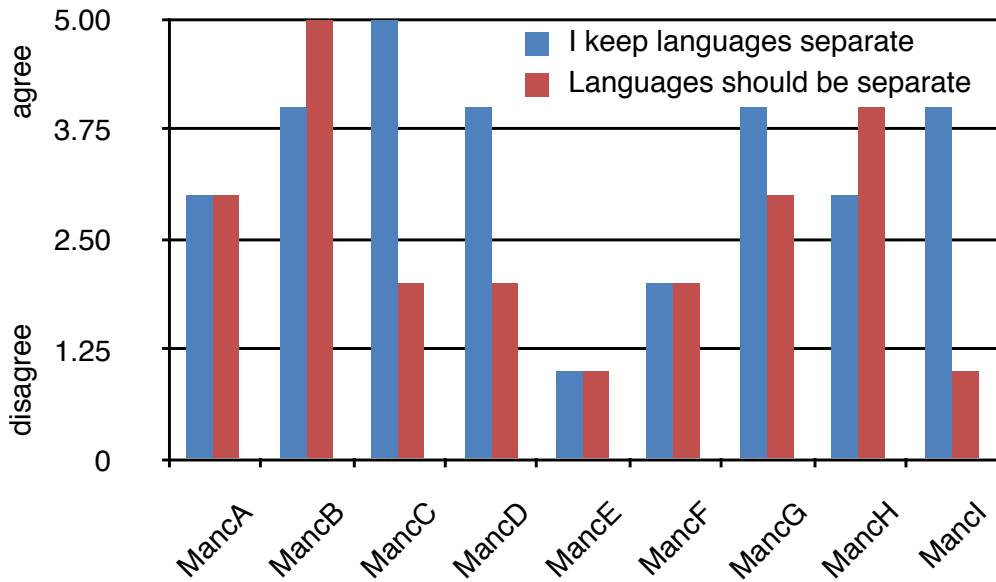




Mean attitudes towards English and Italian

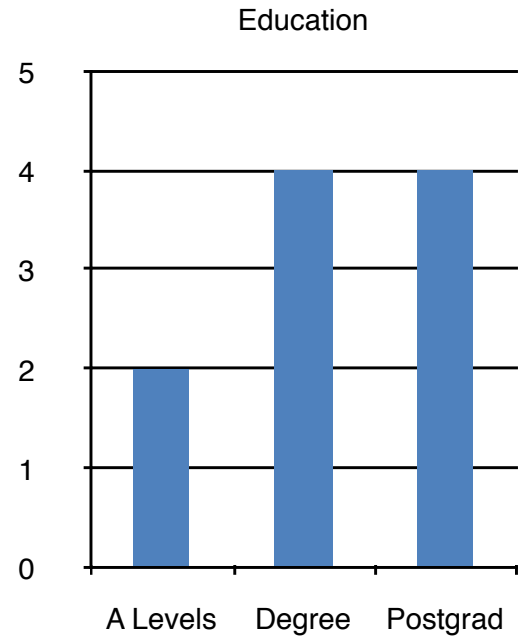
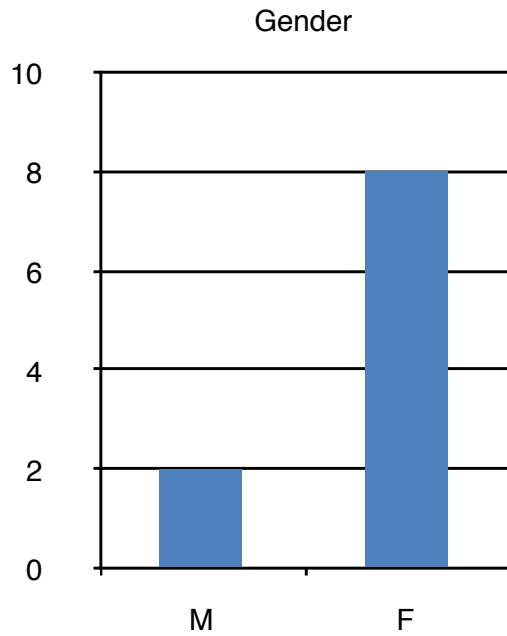


CS self-report and attitude by speaker

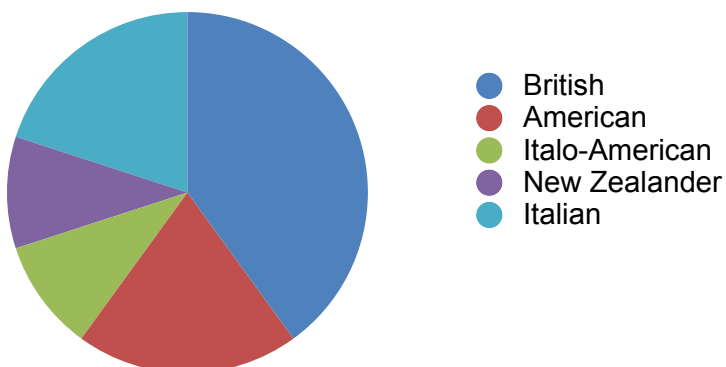


## Appendix 7

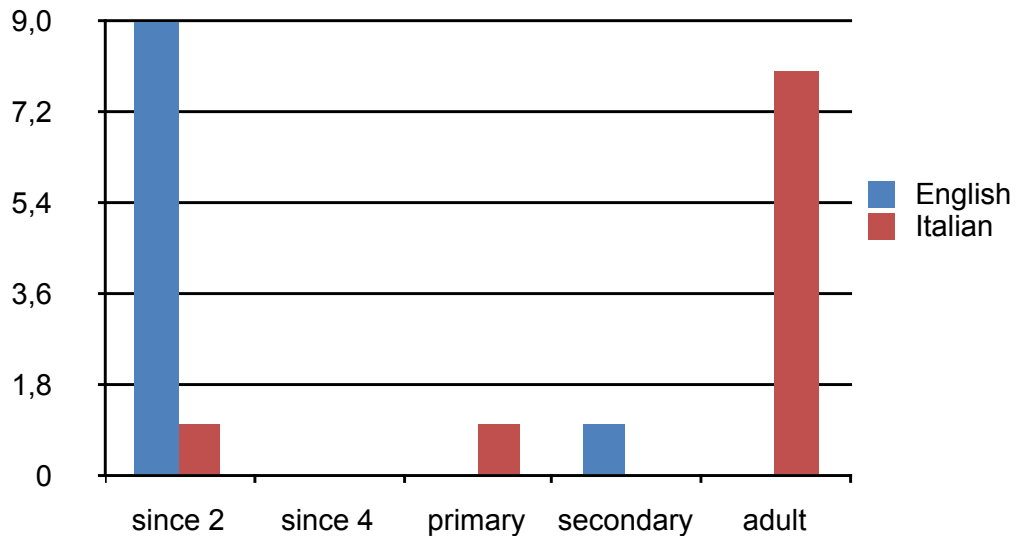
Milan speakers' questionnaire data



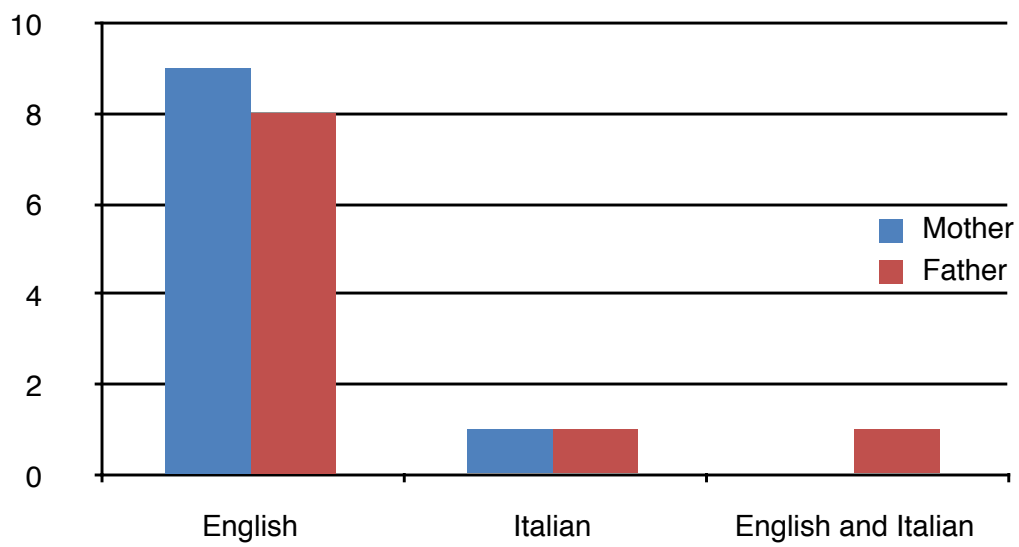
National identity



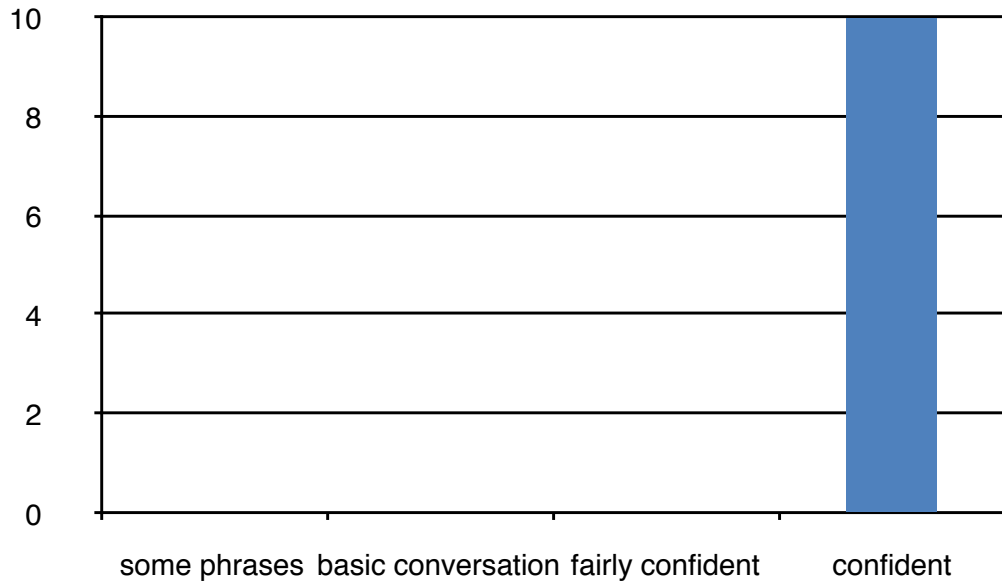
Age of acquisition of English and Italian



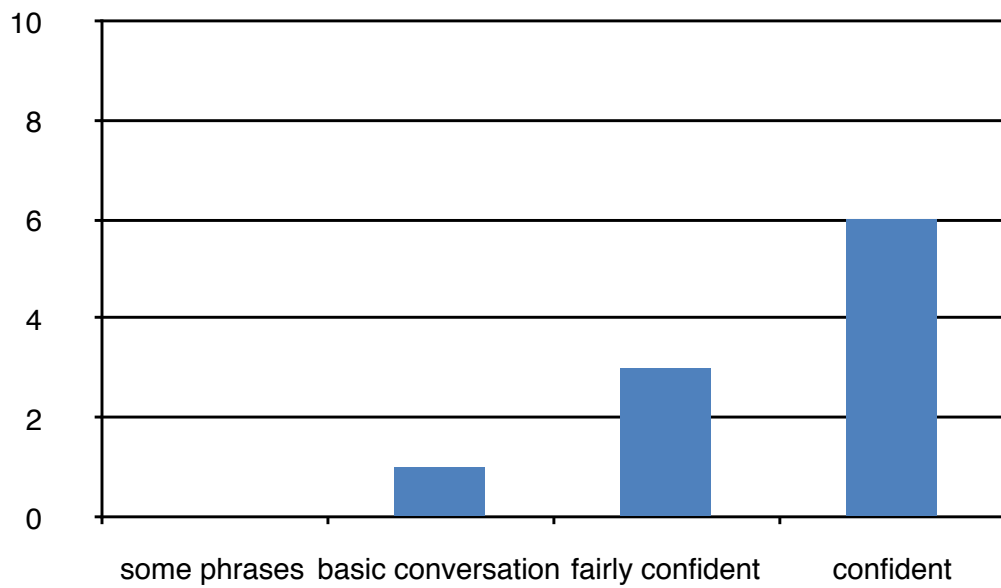
Parental language input



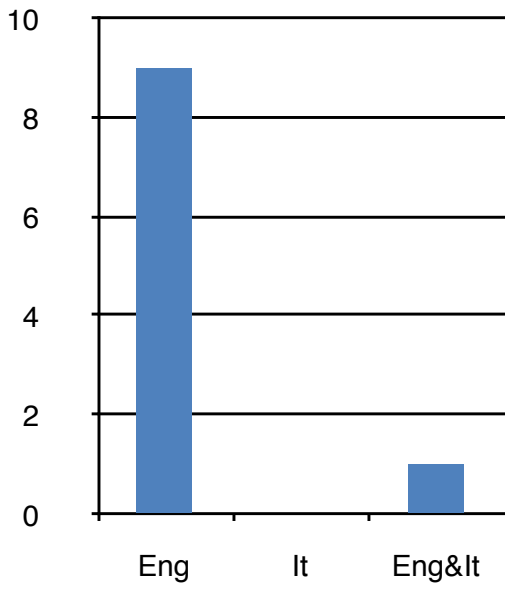
English proficiency



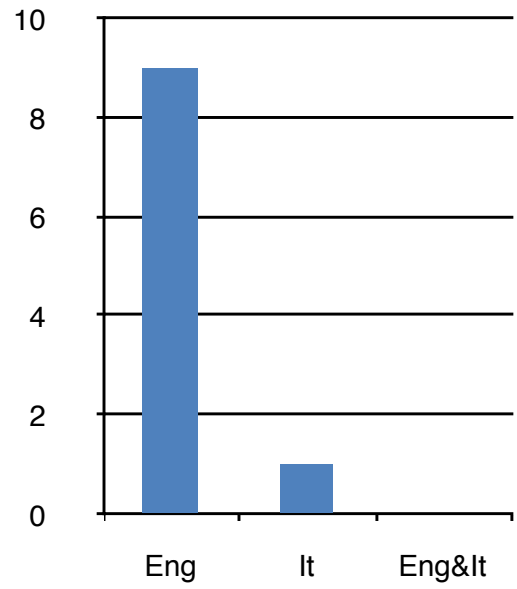
Italian proficiency



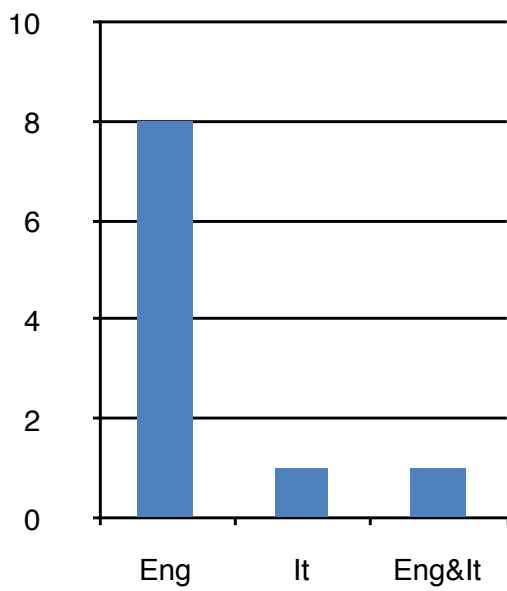
Medium primary school



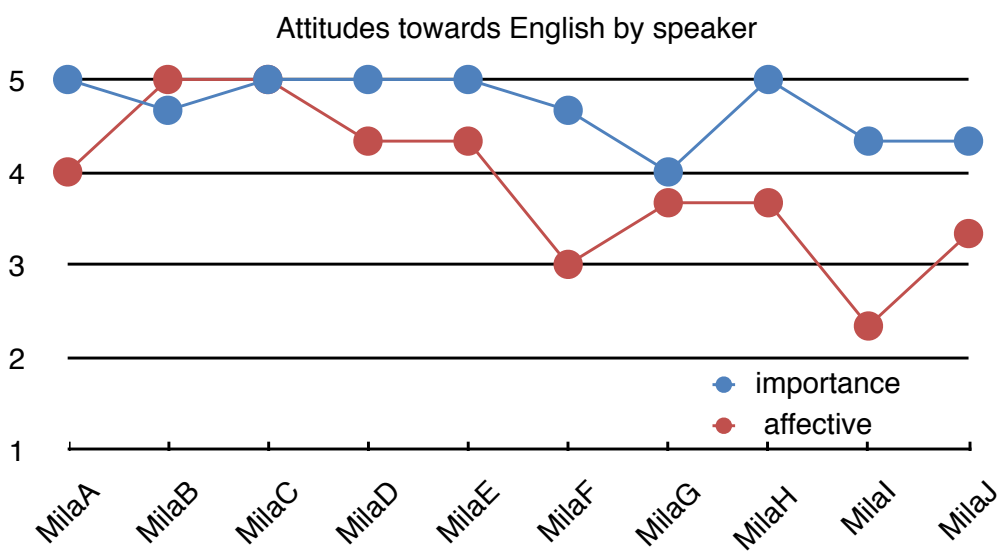
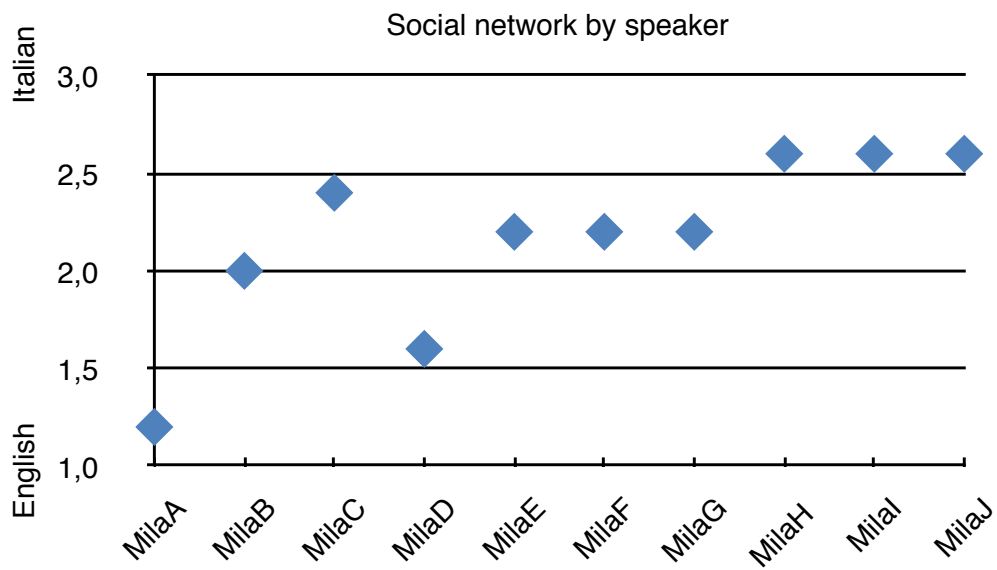
Medium secondary school

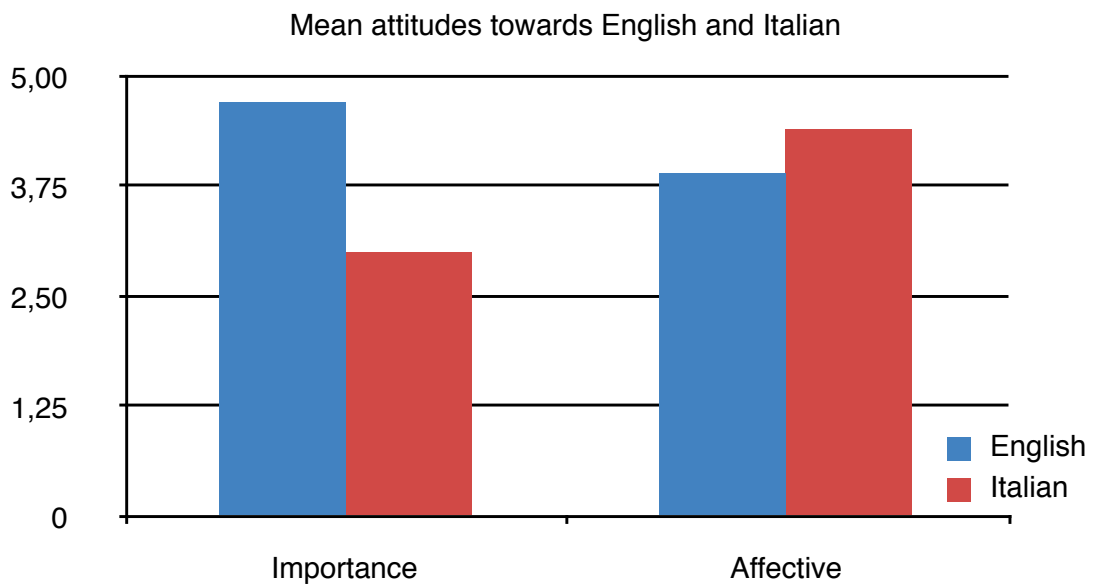
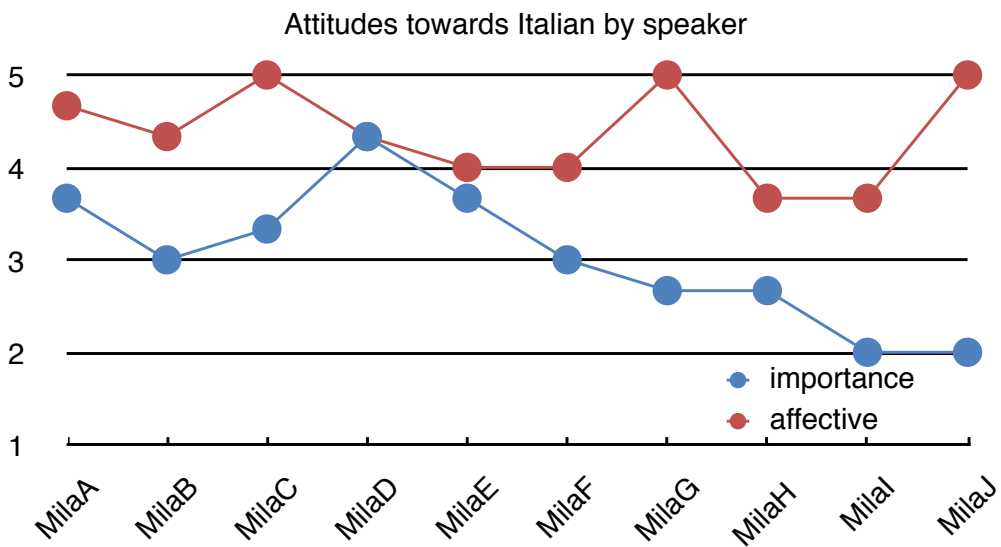


Medium higher education

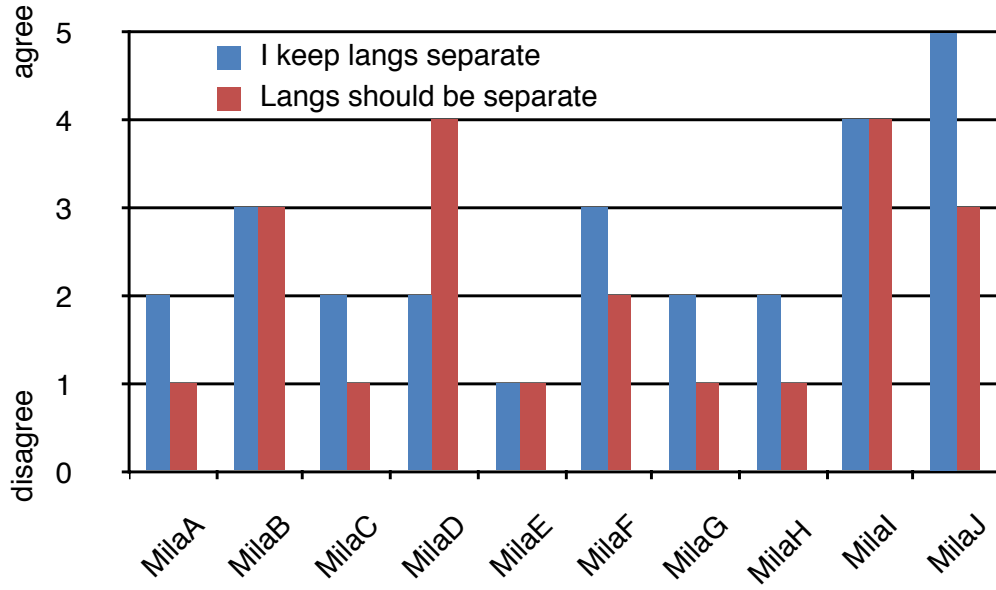








CS self-report and attitudes by speaker



## Appendix 8

Sample of a transcript from the Manchester data.

A sample showing the text of the CHAT transcription of the Manchester 2 recording, including the transcript header.

```
1 @Begin
2 @Languages:      it, en
3 @Participants:   PAO Paolo Adult, LUC Luca Adult
4 @ID:            it, en|bangor4|PAO||male|||Adult||
5 @ID:            it, en|bangor4|LUC||male|||Adult||
6 @Media:         manchester2, sound
7 @Situation:     PAO is a friend of LUC. They are talking in a café at
8                 the University of Manchester .
9 @Date:          14-NOV-2008
10 @Coder:         Alberto Rosignoli, Prifysgol Bangor University
11 @Time Duration: 00:54:21
12 @Comment:       Filename: manchester2.CHA; Soundfile: manchester2.wav
13 @Comment:       Language markers: default = Italian, @en = English, @O =
14                 Undetermined, @s:en+it = word with first morpheme(s) English, second
15                 morpheme(s) Italian, @s:it+en = word with first morpheme(s) Italian,
16                 second morpheme(s) English.
[...]
```

309 \*LUC: invece io ho visto la versione originale .  
310 \*LUC: lei era grandiosa .  
311 \*LUC: anche perché faceva la parte di una vamp@O uh@O del:  
312 Middle@en West@en .  
313 \*LUC: quindi aveva questo accento particolare eccetera .  
314 \*LUC: cioè c' erano molte sottigliezze [>] ### perse .  
315 \*PAO: sì [<] .  
316 \*LUC: è [?] doppiata anche male # da secondo me una doppiatrice  
317 neanche bravissima .  
318 \*LUC: per cui se uno ### italiano vede il film@O dice: +"/.  
319 \*LUC: +” (am)mazza hanno dato l' Oscar@O a: (que)sta:  
320 performance@O ?  
321 \*LUC: cioè molto diversa era in inglese ## indipendentemente dalla qualità  
322 del film@O che non era un granché .  
323 \*LUC: però la performance@O in sé stessa era meravigliosa .  
324 \*PAO: sì .  
325 \*PAO: per esempio # secondo me un: [/] un altro film:@O uhm #  
326 trado(tto) [/] doppiato male # è: Thelma\_and\_  
327 Louise@en .  
328 \*PAO: l' hai mai visto tu in originale ?  
329 \*LUC: no .  
330 \*PAO: allora # ti [//] io adess(o) [>] .  
331 \*LUC: <mi sembrava> [<] doppiato bene perché l' ho visto solo in italiano  
332 [>] .  
333 \*PAO: no [<] ti spiego # perché lì è tutto giocato a livello: di classi  
334 sociali .  
335 \*LUC: mmhm@O .  
336 \*PAO: io adesso non mi ricordo chi era: Thelma@O e chi era

337 Louise@O .  
338 \*PAO: comunque la bionda ## e: [>] .  
339 \*LUC: Gina@O [<] Davis@O [>] # Louise@O .  
340 \*PAO: Gina@O [<] Davis@O .  
341 \*PAO: Louise@O [>] +/.  
342 \*LUC: <no Thelma@O> [<] # Thelma@O .  
343 \*PAO: <era comunque> [//] apparteneva a una classe sociale inferiore .  
344 \*PAO: cioè <nel dopp(iaggio)> [//] <nella uhm@O> [//] nella versione  
345 originale <il suo in(glese)> [/] il suo inglese americano era  
346 orribile .  
347 \*LUC: allora mi sa che è il personaggio di Susan@O Sarandon@O  
348 da quello che mi ricordo .  
349 \*LUC: quella che faceva la cameriera .  
350 \*PAO: sì # quella era # più educated@en .  
351 \*PAO: cioè rispe(tto) [/] # &ri rispetto a: +/.  
352 \*LUC: la cameriera era più educa(ta) [>] +...  
353 \*PAO: sì [<] sì ## perché probabilmente uh@O <Gina@O Davis@O> [//] la: [//]  
354 il personaggio che era: interpretato da Gina@O Davis@O era: #  
355 proprio la campagnola: <da strada eh> [>] .  
356 \*LUC: <ah@O ho capito> [<] .  
357 \*PAO: <e nel: uhm@O> [/] e nel doppiaggio chiaramente # si perde  
358 tutto .  
359 \*LUC: mmhm@O sì [>] perde molto nel doppiaggio anche in:  
360 Sliding\_Doors@O .  
361 \*PAO: sì .  
362 \*LUC: l'hai visto # <Sliding\_Doors@O ## che è molto  
363 giocato> [>] anche quello sull' accento .  
364 \*PAO: <sì però non l' ho visto in inglese> [<] .  
365 \*LUC: è molto bello vederlo in inglese # perché lei parla con un accento  
366 britannico .  
367 \*LUC: lei parla con un accento <londinese bellissimo> [>] .  
368 \*PAO: <posh@en un po'> [<] ?  
369 \*LUC: un po' posh@en .  
370 \*PAO: sì e incontra uno <scozzese # mmhm@O> [>] .  
371 \*LUC: <e incontra questo scozzese> [<] +//.  
372 \*LUC: molto bello # come parlano .  
373 \*LUC: son molto belli <gli accenti in quel film@O> [>] .  
374 \*PAO: <sì sì sì> [<] # mmhm@O .  
375 \*LUC: e lei <ha un:> [/] ha un accento molto: sì molto posh@en molto: ##  
376 proprio da londinese del sud ma molto: [/] molto snob@O così .  
377 \*LUC: <quindi è> [>] molto bello vederlo in inglese .  
378 \*PAO: sì [<] .  
379 \*PAO: sì sì [>] .  
380 \*LUC: cioè [?] [<] è più colorito .  
381 \*PAO: <un al(tro)> [//] un altro: +...  
382 \*PAO: <hai presente> [>] la Tata ?  
383 %com: 'Tata' (Nanny) is the name of the TV series, therefore it is  
384 capitalised here.  
385 \*LUC: sì [<] .  
386 \*LUC: sì .  
387 \*PAO: la tv [//] la: [>] eh .  
388 \*LUC: sì .  
389 \*PAO: <tu lo> [/] tu lo sai che in italiano: +...  
390 \*PAO: probabilmente lo saprai .  
391 \*PAO: <in italiano è complet(amente) e:> [//] cioè in lingua originale # è  
392 completamente diverso .

393 \*LUC: io so che lei è ebrea # in uh@O lingua originale [>] .  
394 \*PAO: sì [<] .  
395 \*LUC: e # la [/] la zia Assunta@O è la madre [>] .  
396 \*PAO: <la madre> [<] sì .  
397 \*PAO: ed è stato # doppiato in questo modo per: # adattarlo a un pubblico:  
398 italiano .  
399 \*LUC: mmhm@O .  
400 \*PAO: nel senso che la Tata comunque va in giro da: circa quindic' anni .  
401 \*PAO: cioè non devi pensare al pubblico italiano di adesso .  
402 \*PAO: devi pensare al pubblico italiano di fine anni ottanta inizio anni  
403 novanta [>] .  
404 \*LUC: mmhm@O [<] .  
405 \*LUC: però non ho capito perché hanno trasformato la madre nella zia .  
406 \*LUC: non potevano renderlo in mamma Assunta@O ?  
407 \*PAO: perché <tra la madre> [///] comunque tra la [/] il personaggio di  
408 zia Assunta@O e +//.  
409 \*PAO: questo lo so perché <l' ho lett(o)> [///] ho <letto varie> [>] +//.  
410 \*LUC: mmhm@O .  
411 \*PAO: +, e il personaggio di zia Assunta@O e lei c' è comunque un:  
412 uhm:@O # un' intesa che e: # non c' è tra una madre e una figlia in  
413 Italia .  
414 \*LUC: ahhah:@O .  
415 \*PAO: cioè non: uhm@O # è abbastanza +//.  
416 \*PAO: e: comunque è un telefilm@O # per una cultura # anglofona o  
417 comunque +//.  
418 \*LUC: sì .  
419 \*PAO: uhm:@O ed è stato adattato .  
420 \*PAO: xx è stato manipolato per: sì sì .  
421 \*LUC: interessante (que)sta cosa .  
422 \*PAO: sì sì .  
423 \*LUC: vedi ci sono tante: [/] tante cose di cui parlare anche in ambiente  
424 filmico in <ambiente cinematografico> [>1] # in ambiente [>2] di  
425 telefilm@O .  
426 \*PAO: sì [<1] .  
427 \*PAO: &e [<2] +//.  
428 \*LUC: potresti fare (que)sto passaggio stavolta # dal letterario alla  
429 mainstream@en # culture@en .  
430 \*PAO: sì sì .  
431 \*PAO: per esempio tornando alla Tata c'è una scena # in  
432 cui ehm c'è la [///] il compleanno della madre .  
433 \*PAO: e: [/] e praticamente nella versione italiana .  
434 \*PAO: cioè quella puntata non ha molto senso perché  
435 favedere # lei che prepara i festoni: uh@O +//.  
436 \*PAO: +” happy@en birthday@en mama@en .  
437 \*PAO: e arriva la zia .  
438 \*PAO: e [/] e nella versione italiana c'è un uhm@O ## voiceover@en .  
439 \*PAO: come si chiama .  
440 \*PAO: mo' non mi viene in italiano il: [>] .  
441 \*LUC: sì [<] .  
442 \*PAO: una [/] una voce: fuori campo [>] # che [/] <che  
443 dic(e)> [///] de(lla) [/] della Tata che dice +//.  
444 \*LUC: eh [<] .  
445 \*PAO: +” festeggio te invece di mia madre perché mia  
446 madre è a Latina@O # in Italia .  
447 \*PAO: +” e tu sei qui .  
448 \*PAO: +” quindi tu &f [///] prendi la: +//.

449 \*PAO: però è stato proprio ## adattato o comunque  
450 manipolato fortemen(te) .  
451 \*PAO: un # infatti è una puntata che non si capisce  
452 niente .  
453 \*PAO: cioè ci sta questa zia .  
454 \*PAO: e tutti dicono +”/ .  
455 \*PAO: +” auguri perché sei la mamma di Francesca@O .  
456 \*PAO: anche il nome è un po’: .  
457 \*PAO: Fran@O si chiama la: [>] in originale quindi .  
458 \*LUC: sì [<] .  
459 \*PAO: no è: interessante: ## è interessante .  
460 \*PAO: <e il> [/] e il per esempio tu l’ hai visto il .  
461 \*LUC: quindi Ietta@O è la nonna .  
462 \*PAO: Ietta@O è la nonna .  
463 \*LUC: ah@O e invece lì nell’ originale chi è Ietta@O .  
464 \*LUC: un’ amica di famiglia forse [>] .  
465 \*PAO: no [<] Ietta è [//] dovrebbe essere la: [/] la  
466 suocera # <la qualche cosa> [>1] sì sì [>2] .  
467 \*LUC: <la suo(cera) eh@O> [<1] .  
468 \*LUC: invece [<2] Ietta@O è la nonna [>] .  
469 \*PAO: è [<] la nonna sì .  
470 \*LUC: Ietta@O [=! laughs] .  
471 \*PAO: &=laughs sì .  
472 \*LUC: che poi è bello anche vedere il nome Ietta@O .  
473 \*LUC: bisognerebbe [>1] vedere il nome: originale <com’ è> [>2] .  
474 \*PAO: sì [<1] .  
475 \*PAO: <sì è> [<2] vero .  
476 \*LUC: Ietta@O [>] .  
477 \*PAO: <perché zia Assunta@O> [<] .  
478 \*PAO: <France(sca)@O xxx> [>] .  
479 \*LUC: <eh@O si chiamerà:> [<] con .  
480 \*LUC: chissà come si chiamerà .  
481 \*LUC: l’ hai mai visto in originale .  
482 \*PAO: sì però non me lo ricordo .  
483 \*LUC: non ti ricordi i titoli [>] .  
484 \*PAO: no [<] # <no no> [>] .  
485 \*LUC: <di [///] cioè i> [<] nomi .  
486 \*PAO: no no .  
487 \*PAO: che poi anche lei è ebrea in realtà .  
488 \*PAO: e: infatti in alcune puntate lei fa queste  
489 celebrazioni particolari: ## della religione ebrea .  
490 \*PAO: <quindi non:> [//] # però non si capisce .  
491 \*PAO: cioè porta il: [/] il cappello [>] tipico .  
492 \*LUC: mmhm@O [<] .  
493 \*PAO: e: [/] e quindi non si capisce non +...  
494 \*PAO: mmhm@O <cioè alcune puntate> [>] sono abbastanza: mmhm@O  
495 .  
496 \*LUC: <beh certo> [<] .  
497 \*PAO: non le capivo .  
498 \*PAO: poi ## sapendo il: fatto che in lingua originale  
499 fosse così: +...  
500 \*LUC: eh@O beh allora ha molto più senso la cosa .  
501 \*PAO: sì sì sì .  
502 \*LUC: ha molto più senso .  
503 \*LUC: quindi potre(sti) [///] lei [///] Hane@O cos’ è che ti  
504 ha detto ?

505 \*PAO: Hane@O mi ha detto in pratica che # e: <devo> [//] uh@O  
506 posso provare a mandare a delle ehm ## e: come  
507 si dice delle ## degli articoli ## giocando sui  
508 problemi sociali socioculturali che possono: # venir  
509 fuori dalla traduzione: eh # <a questi:> [/] <a questi:> [//] a  
510 queste riviste .  
511 \*PAO: New\_Voices@O # conosci ?  
512 \*LUC: no@O .  
513 \*PAO: no .  
514 \*PAO: e comunque <è una rivista:> [//] da come mi ha  
515 detto è una rivista online@O # che si occupa di  
516 queste cose per chi non ha mai pubblicato .  
517 \*LUC: mmhm@O .  
518 \*PAO: e: e poi Mita@O [//] Meta@O # che è canadese .  
519 \*PAO: me l'aveva detto pure Sara@O .  
520 \*PAO: ehm: .  
521 \*LUC: ma tu hai: in mente qualcosa ?  
522 \*LUC: effettivamente hai pensato [>] ?  
523 \*PAO: <per un postdoc@en> [<] ?  
524 \*LUC: sì [<] .  
525 \*PAO: allora # ti spiego .  
526 \*PAO: io avevo pensato uh:@O siccome questo Mister@O Alfred@O xx fa  
527 parte di una trilogia ### e: di tradurre alt(ri) [//]  
528 questi altri due romanzi ### e di # trovare <un &s>  
529 [//] <un certo: non lo so> [//] un certo pattern@en  
530 di: +...  
531 \*PAO: perché questi altri due romanzi # sono simili a  
532 livello di tematica .  
533 \*LUC: mmhm@O .  
534 \*PAO: sempre Glasgow@O degli anni Sessanta: e: # queste  
535 figure di insegnanti che combattono con questa  
536 società e: descolarizzata o quasi descolarizzata ehm@O .  
537 \*PAO: e cioè cercare comunque di: [/] di vedere uno  
538 [/] uno [/] una: [///] di tradurre gli altri due # e  
539 vedere: di trovare un certo: tessuto comune a: +...  
540 \*PAO: però è una cosa che # devo chiedere anche alla  
541 mia: supervisor@en italiana .  
542 \*PAO: ma ## secondo me non risponderà eh@O .  
543 \*LUC: mmhm@O .  
544 \*PAO: sì .  
545 \*PAO: tu che pensi ?  
546 \*LUC: mah potrebbe essere interessante .  
547 \*LUC: cioè è una continuazione comunque di <una ricerca>  
548 [>] iniziata .  
549 \*PAO: sì [<] .  
550 \*LUC: e adesso la prosegui .  
551 \*PAO: <sì # sì sì> [=! whispers] .  
552 \*PAO: e comunque volevo anche proporre: la mia traduzione  
553 a una casa editrice [>] .  
554 \*LUC: <eh beh quello> [<] sì !  
555 \*PAO: uhuh@O .  
556 \*PAO: però # tu dirai +”/ .  
557 \*PAO: +” qui in Inghilterra cosa vai a tra(durre) +// .  
558 \*PAO: cioè si interessano dell'italiano ?  
559 \*PAO: tu che dici ?  
560 \*LUC: eh no dovresti farlo in Italia questo [>] .



561 \*PAO: mmhm@O [<] .  
562 \*PAO: eh ma in Italia dove lo faccio ?  
563 \*PAO: quale casa editrice secondo te ?  
564 \*LUC: non ho la minima idea .  
565 \*LUC: provare la Mondadori@O la: [//] # le solite [>] .  
566 \*PAO: mmhm@O [<] .  
567 \*LUC: Feltrinelli@O queste [>] +//.  
568 \*PAO: sì [<] .  
569 \*LUC: ce ne sono xx così .  
570 \*PAO: sì &=sighs sì sì sì .  
571 \*PAO: ehm:@O +...  
572 \*PAO: tu come avresti tradotto per esempio +”/ .  
573 \*PAO: +” sucking@en infant@en .  
574 \*PAO: e quello fucking@en infant@en capisce .  
575 \*LUC: sucking@en infant@en [>] ?  
576 \*PAO: uh@O [<] .  
577 \*LUC: uh:@O [>] .  
578 \*PAO: perché [<] ti spiego la scena .  
579 \*PAO: ci sta # <questo docente> [/] questo professore alle:  
580 superiori # che sta spiegando .  
581 \*PAO: e ci sta questo: ragazzo # che sta ehm:@O # usando  
582 la: [/] la penna # a mo’ di sigaretta .  
583 \*LUC: sì .  
584 \*PAO: e la sta: uhm@O # picchiettando no [>] ?  
585 \*LUC: mmhm@O [<] .  
586 \*PAO: e: # e distrae <gli alt(ri)> [//] <gli altri> [//]  
587 le altre persone .  
588 \*PAO: e lui dice +”/ .  
589 \*PAO: +” e: # everybody@en would@en uh@O know@en you’re@en a@en sucking@  
en  
590 infant@en .  
591 \*PAO: e il bambino capisce +”/ .  
592 \*PAO: +” ahhah@O # you@en you@en said@en fucking@en infant@en .  
593 \*PAO: +” I’ll@en tell@en my@en mum@en .  
594 \*PAO: come avresti tradotto tu # sucking@en infant@en &fant ?  
595 \*LUC: sucking@en S U C +...  
596 \*PAO: sì # sì sì .  
597 \*LUC: sucking@en .  
598 \*LUC: uh:@O [>] +...  
599 \*PAO: <dopo ti> [<] dico come ho tradotto io .  
600 \*LUC: ognuno sa che sei un: +...  
601 \*PAO: cioè io avevo pensato poppante ## perché poppante è  
602 sucking@en infant@en .  
603 \*PAO: però dovevi trovare # un equivalente ## fon(etico)  
604 ehm [>] +//.  
605 \*LUC: fonetico [<] <insieme a:> [>] +...  
606 \*PAO: +, <fonetico simile> [<] .  
607 \*LUC: uh@O xx fucking@en infant@en come l’hai tradotto pottuto [>]  
608 +//.  
609 %com: Luca makes up the word ‘popputo’ to reproduce the ambiguity effect  
610 Paolo is talking about.  
611 \*PAO: allora io ho tradotto fucking@en infant@en +”/ .  
612 \*PAO: e: tutti penseranno che non vali un’ acca .  
613 \*PAO: e fucking@en infant@en # cioè il bambino percepisce +”/ .  
614 \*PAO: +” che vali meno di una cacca .  
615 \*PAO: quindi <acca cacca> [>] .

616 \*LUC: &=laughs [>] .  
617 \*LUC: <oppure tu [//] che tua madre è una vacca>  
618 [=! laughs] .  
619 \*PAO: sì ah@O però uh@O .  
620 \*PAO: cioè e [/] ed è piaciuta molto alla: [//] sia qui  
621 perché l' ho presentata alla [//] # <alla: uhm@O> [//]  
622 (que)st' estate alla conferenza .  
623 \*PAO: la mia professoressa mi aveva suggerito +"/.  
624 \*PAO: +” non vali ehm +//.  
625 \*PAO: +” pensi che sia un Pinocchio@O .  
626 \*PAO: e lui capisce finocchio .  
627 \*PAO: però [>] ah@O # tu capisci bene che ha delle #  
628 comunque delle: ehm +...  
629 \*LUC: mmhm@O [<] .  
630 \*PAO: cioè il [///] come si dice ### delle: conseguenze  
631 diverse .  
632 \*PAO: cioè <che non:> [///] fucking@en infant@en non c' entra  
633 niente con l' omosessualità cioè .  
634 \*LUC: mmhm@O .  
635 \*PAO: quindi sarei andato un po' troppo fuori strada .  
636 \*PAO: e quindi (que)sta cosa +...  
637 \*PAO: <ma è pieno di (que)sta roba> [>] !  
638 \*LUC: <è [/] è [/] è un gran> [<] lavoro !  
639 \*LUC: è interessante come: [>] +...  
640 \*PAO: sì .  
641 \*LUC: è stato: # un gran lavoro di ingegno di [>] +...  
642 \*PAO: eh [<] sì # &=sighs considerando che l' ho fatto in  
643 un anno e mezzo .  
644 \*LUC: un anno e mezzo ?  
645 \*LUC: non nello standard@O tre anni ?  
646 \*LUC: come mai così poco [>] ?  
647 \*PAO: ah:@O in Italia funziona così ## perché in Italia mi  
648 [/] mi hanno dato prima una: [/] una tesi .  
649 \*PAO: cioè un titolo hanno preso dal cilindro +"/.  
650 \*PAO: +” tieni ## le [/] le traduzioni del: [/] dei  
651 tradu(ttori) [>] +//.  
652 \*LUC: <ah@O perché> [<] non decidi tu in Italia il tuo  
653 dottorato ?  
654 \*PAO: no ## ma scherzi ?  
655 \*PAO: ma nemmeno questo l' ho deciso io !  
656 \*PAO: meno male che mi è piaciuto .  
657 \*LUC: ma in Italia non sei tu che proponi la tua  
658 ricerca ?  
659 \*PAO: no .  
660 \*PAO: o comunque deve piacere al tuo tutor@O [//] # al  
661 tuo: supervi(sor)@en [>] .  
662 \*LUC: <e perché> [<] questa cosa ?  
663 \*LUC: perché poi se la pubblica sotto il suo nome ?  
664 \*PAO: eh .  
665 \*LUC: se la è pubblicata sotto il tuo <nome la tua>  
666 [>] ?  
667 \*PAO: <mmhm:@O no> [<] # che io sappia no .  
668 \*PAO: ma perché sto qua cioè &=laughs .  
669 \*PAO: non perché mi piaccia il tempo voglio dire .  
670 \*LUC: cioè non: sei libero di decidere della tua +//.  
671 \*PAO: <molto influenza(to)> [>] +/.

672 \*LUC: <cioè il: dipartimento> [<] # progetta tot [>] +//.  
673 \*PAO: ma [<] <io avevo> [//] # addirittura io avevo pensato  
674 alla creazione <di un> [/] # di un glossario medico #  
675 perché io all' inizio ero un traduttore medico #  
676 medico italiano eh dal [/] ### dal &set +//.  
677 \*PAO: cioè prendere tipo un [/] un centinatio di parole ##  
678 di termini .  
679 \*PAO: e: e vedere come si sono evolute dal Settecento  
680 ad oggi .  
681 \*PAO: siccome lì nessuno era tecnico # nessuno era in  
682 grado di seguirmi # m' han detto +"/.  
683 \*PAO: +” no .  
684 \*PAO: +” fai Darwin@O !  
685 \*PAO: son stato # due anni su Darwin@O .  
686 \*PAO: e ho fatto +"/.  
687 \*PAO: +” che faccio ?  
688 \*PAO: ho cambiato .  
689 \*PAO: e: &m: la prof@O mi ha dato: +"/.  
690 \*PAO: +” ti interessa la: [/] # la: [/] la letteratura  
691 scozzese ?  
692 \*PAO: e io ho fatto +"/.  
693 \*PAO: +” no .  
694 \*PAO: però: mi ha proposto questo libro pieno di giochi  
695 di parole .  
696 \*PAO: <mi ha:> [/] mi ha appassionato .  
697 \*PAO: ed eccomi qua .  
698 \*LUC: ma roba da matti !  
699 \*LUC: e <ti credo> [>] che la ricerca va a bagno se &s  
700 non si fa neanche una +...  
701 \*PAO: sì [<] .  
702 \*LUC: cioè devi fare una <cosa che interessa> [>1] al  
703 <dipartimento xx> [>2] +/.  
704 \*PAO: <a &tu> [<1] +//.  
705 \*PAO: <xx certo> [<] !  
706 \*LUC: no qui in Inghilterra sei tu che proponi .  
707 \*LUC: <se interessa poi> [>] al dipartimento bene .  
708 \*PAO: <sì eh> [<] .  
709 \*LUC: <sei tu che fai la tua> [>] research@en proposal@en .  
710 \*PAO: <cioè se qui> [<] +//.  
711 \*LUC: quindi la research@en proposal@en non esiste in Italia  
712 non +...  
713 \*PAO: è molto influenzabi(1e) +/.  
714 \*PAO: cioè ### sì sì sì non +...  
715 \*PAO: e tu invece ehm:@O qual è la: [/] <la tua tesi>  
716 [//] la [//] il tuo argomento: +//.  
717 \*PAO: Dario@O Argento@O # cioè come [/] come la pittura  
718 influisce su: +...  
719 \*PAO: quindi hai fatto pure il film:@O la: &k la: #  
720 Sindrome di <Stendhal@O # mmhm> [>] .  
721 \*LUC: <eh beh sì # ci> [<] sarà tutto un capitolo  
722 dedicato a quello .  
723 \*PAO: t' è piaciuto # il film@O ?  
724 \*LUC: uh@O l' ho rivalutato molto # facendo (que)sti studi .  
725 \*LUC: a: <primo impatto> [>] no .  
726 \*PAO: mmhm@O [<] .  
727 \*PAO: perché no ?

728 \*LUC: ehm: l' ho trovato molto banale come storia a &li  
729 come a livello di intreccio così .  
730 \*LUC: però ### l' ho rivalutato molto poi facendo: [//]  
731 uhm@O facendo questi studi apprezzando tutti i  
732 riferimenti e: culturali e iconografici che ha:  
733 diciamo ## inserito all' interno del film@O .  
734 \*LUC: allora: l' ho parecchio rivalutato .  
735 \*LUC: e anche certi aspetti +//.  
736 \*LUC: e anche il Fantasma dell' Opera che l' avevo  
737 odiato adesso # ho rivalutato certi aspetti  
738 soprattutto a livello di utilizzo di una luce  
739 particolare eccetera .  
740 \*LUC: insomma ## non: rimangono tra i miei film@O preferiti  
741 .  
742 \*LUC: però li ho rivalutati da # uh@O alcuni punti di  
743 vista &=yawns .  
744 \*LUC: io &=coughs discuto nella mia tesi l' autorialità di  
745 Dario@O Argento@O perché # c' è una diatriba enorme  
746 sul fatto +//.  
747 \*LUC: dato che lui proviene da dei generi popolari che  
748 sono l' horror@O il giallo eccetera # mmhm@O +...  
749 \*LUC: +, nella [//] diciamo nella critica <degli anni  
750 novanta in> [//] dagli anni novanta in poi hanno #  
751 iniziato a chiamarlo autore .  
752 \*LUC: è stato premiato dalla Cinematheque@O eccetera e qui  
753 e là .  
754 \*LUC: allora io ho individuato come +//.  
755 \*LUC: mmhm@O dato che è stato studiato molto nel suo  
756 stile qui e là .  
757 \*LUC: io ho uhm:@O voluto individuare nel eh riferimenti  
758 alla pittura # <come suo marchio di> [//] come uno  
759 dei suoi marchi di autorialità # i suoi #  
760 incredibili rimandi .  
761 \*LUC: e oltretutto voglio analizzare come uhm:@O ### i vari  
762 rimandi alla pittura eccetera <possono essere> [//]  
763 possano essere inglobati # eh nella percezione nella  
764 cultura del gotico # uh@O eccetera .  
765 \*LUC: quindi ehm: do un' interpretazione uh@O in chiave uh@O  
766 allegorica eccetera ai ai ai ai ai dipinti  
767 eccetera che vengono # uh:@O referenced@en # <uhm@O no da  
768 un punto di vista appunto:> [///] <uhm@O come dire  
769 da un> [///] dandoci una chiave di lettura del:  
770 [///] gotica # eccetera .  
771 \*LUC: <quindi all' interno> [=! whispers] +...  
772 \*PAO: e ti trovi: bene .  
773 \*PAO: cioè hai scritto il primo capitolo quindi +...  
774 \*LUC: no non ho scritto .  
775 \*LUC: ci sono dietro a scrivere il primo capitolo .  
776 \*LUC: consegnerò a febbraio .  
777 \*PAO: ahha@O .  
778 \*LUC: ho scritto una ventina di pagine .  
779 \*LUC: il primo capitolo dovrà essere un trentacinque  
780 quaranta pagine .  
781 \*PAO: eh mah e quando hai iniziato il dottorato l'  
782 anno scorso [>] ?  
783 \*LUC: <l' anno> [<] scorso # ottobre scorso .

784 \*PAO: cioè tu praticamente sei venuto in Inghilterra .  
785 \*PAO: hai fatto prima il lettore d' italiano .  
786 \*LUC: sì # due anni .  
787 \*PAO: due anni .  
788 \*PAO: e poi hai fatto questa: proposal@en .  
789 \*PAO: l' hai fatta direttamente qui o a tutti cioè:  
790 [>] +/.  
791 \*LUC: no [<] no no no l' ho fatta: a cinque posti .  
792 \*PAO: ma perché sapevi che cercavano o &ela +...  
793 \*PAO: qua la proposal@en come si fa cioè [>] +...  
794 \*LUC: allora uh: son stato motivato da: ## ragioni diverse  
795 .  
796 \*LUC: cioè ## <quando allora> [///] innanzitutto uh@O l' ho  
797 mandato in cinque: università diverse .  
798 \*LUC: tre per motivazione di professore .  
799 \*LUC: due per motivazione di: borsa di studio .  
800 \*PAO: ah@O .  
801 \*LUC: cioè # uh:@O ## tre a livello di professore perché  
802 <c' erano> [//] in queste università c' erano degli  
803 esperti nel: film@O di genere # quindi c' erano  
804 degli esperti nel film@O gialli nell' horror@O # <e in  
805 certi> [///] italiano eccetera .  
806 \*LUC: quindi mi potevano aiutare dal punto di vista uh:  
807 [>] .  
808 \*PAO: <mi stai> [<] facendo venire in mente una: [//] #  
809 un' idea .  
810 \*LUC: sì sì tu <butta giù> [>] !  
811 \*PAO: <vai vai> [<] .  
812 \*LUC: e invece due tra cui Manchester@O xxx motivi che c'  
813 era una: [/] una borsa di studio ## parecchio buona  
814 che era qui e a Birmingham@O .  
815 \*PAO: mmhm@O .  
816 \*LUC: alla fine ## mi avevano accettato in altri posti .  
817 \*LUC: ma visto che la borsa di studio era quasi  
818 assente ## o copriva solo le tuition@en fees@en ## allora  
819 ho: uhm@O # cambiato .  
820 \*LUC: e sono passato a Manchester@O perché comunque: avevo  
821 la possibilità di mantenermi eccetera .  
822 \*PAO: mmhm@O .  
823 \*LUC: mmhm@O .  
824 \*PAO: e dove ehm:@O +//.  
825 \*PAO: all' inizio stavi a Nottingham@O no ?  
826 \*LUC: no # a Sheffield@O .  
827 \*PAO: Sheffield@O .  
828 \*LUC: mmhm@O # un incubo .  
829 \*PAO: eh ?  
830 \*LUC: un incubo .  
831 \*PAO: perché ?  
832 \*LUC: oh@O noiosa come città è una menata: !  
833 \*LUC: che idea ti è venuta ?  
834 \*PAO: no gli esperti di puns@en and@en wordplay@en .  
835 \*PAO: cioè io devo mandare anche # e: queste: [//] questa  
836 proposta # a qualcuno che <fa &es> [///] <&pa parla>  
837 [///] comunque tratta di traduzione di puns@en .  
838 \*LUC: ahha@O sì !  
839 \*PAO: conosci qualcuno ?

840 \*LUC: no .  
841 \*PAO: mmhm@0 [>] .  
842 \*LUC: <non è> [<] <il mio:> [/] il mio genere .  
843 \*LUC: dovresti guardare ## nel dipartimento d' Inglese qui  
844 se c' è qualcuno che: gioca su questi +...  
845 \*PAO: beh però è +/.  
846 \*LUC: beh &sans intendi come nell' opera di: # uhm:@0 <che  
847 so io uhm@0> [//] # <come si chiama> [//] Oscar@0 Wilde@0  
848 .  
849 \*LUC: che è the@en Importance@en to@en be@en Earnest@en # che Earnest è  
850 il nome .  
851 \*LUC: ma earnest@en è anche nel senso di honest@en .  
852 \*LUC: cioè questi <sono i pun@en> [>] ?  
853 \*PAO: <sì sì sì> [<] <i giochi di parole> [>] .  
854 \*LUC: <i giochi di parola> [<] che in Italia è l'  
855 importanza di chiamarsi Ernesto # e che vuol dire  
856 anche l' importanza di essere onesto .  
857 \*LUC: ma <da: uhm:@0> [//] da noi si perde .  
858 \*PAO: sì # questo è .  
859 \*PAO: il problema è che questo libro qui non dà # solo  
860 questi spunti # ma dà anche problemi di natura  
861 sociale e comunque: +...  
862 \*PAO: &=coughs cioè la differenza tra inglese standard@0 e  
863 inglese colloquiale .  
864 \*PAO: e che l' inglese colloquiale in realtà # <è: uh:@0>  
865 [//] è espressione dell' animo del personaggio .  
866 \*PAO: anche # comunque è vero .  
867 \*PAO: cioè nel senso che: # gli inglesi sono molto  
868 classisti .  
869 \*PAO: cioè ogni classe <# sociale ha> [>] il suo inglese .  
870  
871 \*LUC: <sì sì> [<] .  
872 \*PAO: quindi <attraverso # lo studio> [>] della lingua si  
873 risale anche allo stato sociale .

## Appendix 9

Sample of a transcript from the Milan data.

A sample showing the text of the CHAT transcription of the Milano 2 recording, including the transcript header.

```
1 @Begin
2 @Languages:      eng, ita
3 @Participants:   EMM Emma Adult, KEL Kelly Adult
4 @ID:             eng, ita|bangor4|EMM||female||Adult|
5 @ID:             eng, ita|bangor4|EMM||female||Adult|
6 @Situation:      EMM and KEL used to be colleagues until a few weeks before
7                  the recording. They are talking in what used to be their office (EMM
8                  is still working there at the time of the recording).
9 @Date:           23-APR-2009
10 @Coder:          Alberto Rosignoli, Prifysgol Bangor University
11 @Time Duration: 00:49:08
12 @Comment:        Filename: milano2.CHA; Soundfile: milano2.wav
13 @Comment:        Language markers: default = English, @it = Italian @0 =
14                  Undetermined, @en+it = word with first morpheme(s) English, second
15                  morpheme(s) Italian, @it+en = word with first morpheme(s) Italian,
16                  second morpheme(s) .
17 [...]
193 *EMM: <really # <did she> [/] uh@0 did> [<] she also have
194    a baby@0 your size [>] .
195 *KEL: <yeah (be)cause my> [<] cousin Mark@0 is only six
196    months older than I am .
197 *KEL: and so when I was first born +...
198 *EMM: ah@0 that's <the &v> [//] the very clever one .
199 *KEL: yes # very <very clever> [>1] # very good-looking and
200    clever [>2] .
201 *EMM: mmhm@0 [<1] .
202 *EMM: mmhm@0 [<2] .
203 *EMM: the furbo@it .
204 *KEL: yes .
205 *KEL: <he's a> [/] he's a good boy .
206 *KEL: anyway .
207 *KEL: so he's [//] he was still breast feeding when I
208    was born .
209 *KEL: <and as long as> [>] +...
210 *EMM: ahhah@0 [<] .
211 *KEL: oh@0 I have a creepy story to tell you .
212 *KEL: anyway but as long as uhm@0 ## the [//] there is
213    this need for milk your body will keep <producing
214    it> [>] .
215 *EMM: <yes of> [<] course .
216 *KEL: so: [>] +...
217 *EMM: <so she was> [<] your wet nurse [>] .
218 *KEL: yeah [<] because my mum would do the pump thing
219    and put it in a bottle .
220 *EMM: mmhm@0 [>] .
```

221 \*KEL: <but then maybe> [<] I'd eat it all because  
 222 apparently I was a piggy .  
 223 \*EMM: yes .  
 224 \*KEL: and uhm:@0 I'd drink all my milk .  
 225 \*KEL: and then I'd cry .  
 226 \*KEL: and my aunt would say [>] +"/.  
 227 \*EMM: aw@0 [<] .  
 228 \*KEL: +" well .  
 229 \*KEL: and so she'd have two babies on [//] # one on each #  
 230 boob .  
 231 \*KEL: and: # anyway that was [>] +...  
 232 \*EMM: scary [<] !  
 233 \*KEL: anyways Fabrizio@0 was saying something like +"/.  
 234 \*KEL: I don't remember .  
 235 \*KEL: but [///] the [///] I [///] I'm convinced that I have  
 236 to tell you this freaky story .  
 237 \*KEL: so we went on Monday to see this baby@0 .  
 238 \*KEL: and we <were talking> [>] about +...  
 239 \*EMM: <oh@0 yes> [<] .  
 240 \*KEL: because on Saturday or Sunday or something like  
 241 that we were at the park .  
 242 \*KEL: and: uhm@0 my friend Sophie@0 who introduced me to  
 243 Fabrizio@0 ### she's got this boyfriend who: is an  
 244 uncle .  
 245 \*KEL: ok@0 <just to get> [>] around all of the: # lineage  
 246 stuff .  
 247 \*EMM: ok@0 [<] .  
 248 \*KEL: anyway and they were playing auntie and uncle .  
 249 \*KEL: and took the: kiddies to the park .  
 250 \*EMM: mmhm@0 .  
 251 \*KEL: and uhm:@0 # the little boy Fausto@0 he's five # or  
 252 six .  
 253 \*KEL: and he just had a new little brother Paolo@0 ##  
 254 who's nine months old .  
 255 \*KEL: and the mum came to meet them at the park with  
 256 Paolo@0 # just to get out of the house (be)cause  
 257 they're moving house .  
 258 \*KEL: and: the little boy is nine months old .  
 259 \*KEL: and she was saying +"/.  
 260 \*KEL: +" oh@0 yeah you know he's still on the boob .  
 261 \*KEL: and I was like +"/.  
 262 \*KEL: +" wow@0 you know it might be time to think about  
 263 weening him # because <at about> [///] you really  
 264 shouldn't breast feed after a year .  
 265 \*KEL: like that's [///] even a year is really pushing it  
 266 .  
 267 \*KEL: at about nine months you should get them off .  
 268 \*KEL: and: [>] .  
 269 \*EMM: &=laughs [<] .  
 270 \*EMM: get them off [>] !  
 271 \*KEL: <and so I> [<] was saying +"/.  
 272 \*KEL: +" yeah <it's kind of> [//] it's a little bit strange but [>] #  
 273 whatever .  
 274 \*EMM: mmhm@0 [<] .  
 275 \*KEL: anyway so then she blurts out that with her  
 276 youngest sister # the mother breast fed until she



277 was three .  
 278 \*EMM: that's not uncommon .  
 279 \*KEL: that is disgusting [>] !  
 280 \*EMM: <it's weird> [<] .  
 281 \*EMM: but it's not uncommon .  
 282 \*EMM: it does happen .  
 283 \*KEL: that's nasty !  
 284 \*KEL: and then she said +"/.  
 285 \*KEL: +" yeah well my mum went a little bit crazy when  
 286 she got pregnant with # Anna@O is the little girl's  
 287 name # because she [///] it was completely unexpected  
 288 .  
 289 \*KEL: she was like fifty .  
 290 \*KEL: <and: there's> [>] <I think like a twelve> [//] ##  
 291 yeah like a twelve year difference between their  
 292 youngest daughter and this accident baby@O .  
 293 \*EMM: wow@O [<] .  
 294 \*KEL: so there's like a: uhm@O like a seventeen year xx  
 295 difference between Audrey@O who's just had a baby  
 296 and ## Anna@O the youngest .  
 297 \*EMM: yes .  
 298 \*KEL: anyway but she went apparently nuts .  
 299 \*KEL: and she breast fed till she was three .  
 300 \*KEL: and I thought that was gross .  
 301 \*KEL: xxx +"/.  
 302 \*KEL: +" ah@O I really stuck my foot in my mouth again .  
 303  
 304 \*EMM: <does this> [///] here's a question .  
 305 \*EMM: do you think that the: [///] if [/] if someone is  
 306 your mentor # are you their mentoree or their  
 307 mentee ?  
 308 %com: neither "mentoree" or "mentee" are listed on the dictionary  
 309 \*KEL: ah@O xx it's your protégé !  
 310 \*EMM: I'm not gonna [: going to] write +"/.  
 311 \*EMM: +" dear protégés of professor@O Lunardi@O .  
 312 \*KEL: oh@O I would say mentorees then [>] .  
 313 \*EMM: mentoree [<] .  
 314 \*EMM: because I'm sure I heard mentees before .  
 315 \*EMM: but dear &m +/.  
 316 \*EMM: I'm writing +/.  
 317 \*EMM: +" dear mentorees .  
 318 \*KEL: also because you have to consider that these  
 319 people [///] most of them aren't native English speakers  
 320 and they'll <understand mentorees> [>] .  
 321 \*EMM: exactly [<] .  
 322 \*EMM: that's why certain changes for: uhm@O Ruggeri@O this  
 323 morning +...  
 324 \*EMM: I just thought +"/.  
 325 \*EMM: +" &=gasps # leave it like it is because they're all  
 326 stupid anyway !  
 327 \*KEL: (be)cause some things they understand # not perfect  
 328 English <a lot better> [>] .  
 329 \*EMM: exactly [<] and so: +...  
 330 \*EMM: that's right .  
 331 \*EMM: there was something that was really stupidly written  
 332 .

333 \*EMM: but I didn't even say anything to him because I  
334 thought +"/.  
335 \*EMM: +" well it's actually quite clear as it is to a  
336 dumbo .  
337 \*EMM: <so probably> [>1] &=chuckles it'll [>2] be ok@O .  
338 \*KEL: mmhm@O [<1] .  
339 \*KEL: yeah [<2] .  
340 \*KEL: that's really &s +//.  
341 \*KEL: oh@O I wanna [: want to] hear this uhm:@O gossip@O about  
342 Bianchi@O .  
343 \*EMM: oh@O yeah # the &t [///] well Stephanie@O Quinn@O # uh@O had  
344 a little tantrum # because she missed the first  
345 two days [>1] # which is <more than eighty> [//]  
346 more than twenty per cent [>2] ### absence .  
347 \*KEL: mmhm@O [<1] .  
348 \*KEL: mmhm@O [<2] .  
349 \*EMM: so she should fail the course .  
350 \*KEL: mmhm@O .  
351 \*EMM: and when he said to her +"/.  
352 \*EMM: +" well # you're very welcome to be here .  
353 \*EMM: +" but I'm not going to assess you .  
354 \*EMM: +" and I'm not going to grade you for being part  
355 of this .  
356 \*EMM: +" and you can take part in your group .  
357 \*EMM: +" and do the whole thi(ng) +//.  
358 \*EMM: +" have the experience .  
359 \*EMM: +" but you're not # going to get a grade .  
360 \*EMM: she freaked out .  
361 \*EMM: and well started shouting .  
362 \*EMM: and say +"/.  
363 \*EMM: +" oh@O I hate this degree .  
364 \*EMM: +" I hate this M\_B\_A@O .  
365 \*EMM: +" it's the +//.  
366 \*EMM: what did she say ?  
367 \*EMM: something like +"/.  
368 \*EMM: +" this is the biggest # expense I've ever had in  
369 my life .  
370 \*EMM: +" and <it's &g> [//] # it sucks kind of thing .  
371 \*EMM: and uhm@O ## then she was complaining about her  
372 colleagues as well .  
373 \*EMM: but I don't know who or what or why .  
374 \*EMM: but <they # xx> [//] it was the professors who told  
375 us .  
376 \*EMM: she's complaining about the others on the course:  
377 and everything .  
378 \*KEL: about how they're not happy with the M\_B\_A@O ?  
379 \*KEL: or about how they suck <as uhm:@O colleagues> [>] ?  
380 \*EMM: <well according to her> [<] # yeah her even yeah  
381 the whole programme is terrible .  
382 \*EMM: she's had &fif [//] fifteen different professors in  
383 the last ## three months .  
384 \*EMM: and we were like +"/.  
385 \*EMM: +" well this is the nature of an international  
386 programme with an international # thingy bob .  
387 \*EMM: and [>] # this is the thing .  
388 \*KEL: mmhm@O [<] .

389 \*EMM: and uhm@O ## she was really upset .  
390 \*EMM: but [///] and saying she [///] it should have been  
391 explained to her that if she missed more than  
392 twenty per cent of the programme she would be  
393 off .  
394 \*EMM: think [>] .  
395 \*KEL: <she knew that in the> [<] beginning .  
396 \*EMM: but she knew that .  
397 \*EMM: and they've all known that .  
398 \*EMM: it's in the rules uhm@O [>] +...  
399 \*KEL: yeah [<] .  
400 \*EMM: the [///] it's the <same for every course they  
401 take> [>] .  
402 \*KEL: <and they can do the math themselves> [<] .  
403 \*EMM: exactly .  
404 \*EMM: <but she wasn't> [>] happy .  
405 \*KEL: that [<] +...  
406 \*KEL: <that for> [///] to me that's ridiculous ## <that she would do> [>]  
407 +...  
408 \*EMM: <and then Bianchi@O is not happy> [<] because all of  
409 these people who are highlighted +...  
410 \*EMM: I need to add a couple more because I've added  
411 them in in pencil but I haven't highlighted them .  
412  
413 \*EMM: uhm@O ## all these people are absent for either one  
414 day or: more than one day .  
415 \*EMM: and if they're absent for one day that's fine .  
416 \*EMM: but more than one day is not fine .  
417 \*EMM: and then those with just dots next to them just  
418 little yellow dots were unjustified in not being  
419 present .  
420 \*KEL: because <we xx> [///] we anticipated that they were going to  
421 be gone because: uhm@O oh@O my gosh # <because they  
422 came in> [>1] to say that they were going to be  
423 gone [>2] .  
424 \*EMM: <there are lots> [<1] .  
425 \*EMM: mmhm@O [<] .  
426 \*KEL: and that they were going to have a problem .  
427 \*KEL: or kind of hinted that they were going to have  
428 a problem # attending class .  
429 \*KEL: but <it's ins(ane)> [///] <it's a &l> [///] that's a lot .  
430 \*KEL: Hermann@O is always gone # always gone .  
431 \*EMM: well it was really funny because I was only  
432 thinking the other day +"/.  
433 \*EMM: +" oh@O it's interesting that she hasn't gone to  
434 Peru@O this time .  
435 \*EMM: +" and that she hasn't told us she's not going  
436 to be here .  
437 \*EMM: and maybe she's being good .  
438 \*EMM: and decided to knuckle down as it's the last  
439 thing [///] # last push before # the summer .  
440 \*EMM: and no # she just didn't turn up .  
441 \*EMM: and she hasn't ## told anybody why .  
442 \*KEL: I would be so happy to flunk so many of these  
443 people though .  
444 \*EMM: that's what Lorenza@O said .

445 \*KEL: because they really never ever show up .  
446 \*KEL: I mean there are some people ## <that uhm:@O> [//] ###  
447 that are good that show up all the time .  
448 \*KEL: but then there are some that really just always  
449 +...  
450 \*KEL: well Hermann@O is really the only one that # stands  
451 out in my mind as somebody who's always gone #  
452 <always gone> [>] .  
453 \*EMM: mmhm@O [<] .  
454 \*KEL: she's <taken so> [>1] many # leaves of absence [>2] .  
455 \*EMM: <yes it's true> [<1] .  
456 \*EMM: yes [<2] .  
457 \*KEL: uhm@O .  
458 @Comment: There is a brief silence.  
459 \*KEL: wow@O ## <so what> [>] .  
460 \*EMM: <so he wants to see> [<] fourteen different &men  
461 little mentorees # for fourteen little half an  
462 hours # on Tuesday: no <Wednesday and Thursday> [>]  
463 .  
464 \*KEL: <how come Quinn@O isn't> [<] highlighted ?  
465 \*EMM: because I didn't know she wasn't there .  
466 \*KEL: ahhah@O [>] .  
467 \*EMM: <only this> [<] morning when he came in and told  
468 me she wasn't there ## <did I know> [>] .  
469 \*KEL: <oh@O (be)cause that's all our way> [<] to check the  
470 uhm@O timbraring@it&een [>] .  
471 \*EMM: <the [//] he's> [<] checking +...  
472 \*EMM: no the timbraring@it&een thing hasn't been checked .  
473 \*EMM: he is going round physically ticking them off [>] #  
474 on his little list .  
475 \*KEL: mmhm@O [<] .  
476 \*KEL: mmhm@O .  
477 \*EMM: these are the ones who write to us .  
478 \*EMM: <or the dots> [>] are the ones he's told us  
479 about .  
480 \*KEL: mmhm@O [<] .  
481 \*EMM: and we don't know # why they're not here .  
482 \*KEL: mmhm@O # mmhm@O .  
483 \*EMM: so he's not happy .  
484 \*EMM: oh@O and the other thing is that # on Tuesday when  
485 all those dot people were not here # and Lunardi@O  
486 asked me +"/.  
487 \*EMM: +" how's it going ?  
488 \*EMM: and I said +"/.  
489 \*EMM: +" oh@O well Bianchi@O wasn't very happy because all  
490 of these people weren't here and he had to +...  
491 \*EMM: +" if you look at the first group there are six  
492 people # there were only two people present .  
493 \*EMM: +" so he had to ## add them # to another group .  
494 \*EMM: and then of course by the next day that group  
495 then became a group of nine people because these  
496 people all showed up .  
497 \*KEL: mmhm@O .  
498 \*EMM: but Lunardi@O just said +"/.  
499 \*EMM: +" oh@O that's not as bad as this is usually .  
500 \*EMM: +" it's normally much worse than that .

501 \*EMM: +" and so <don't worry> [>] .  
502 \*KEL: <that's not acceptable> [<] .  
503 \*EMM: no [>] !  
504 \*KEL: even [<] if it's <usually much worse it's never  
505 acceptable> [>] .  
506 \*EMM: +" <oh@0 it's only five people> [<] .  
507 \*EMM: yeah .  
508 \*EMM: +" it's only a handful .  
509 \*EMM: he said +" .  
510 \*EMM: meanwhile the professor was like +...  
511 \*KEL: I think it's really funny though that Quinn@0 blew  
512 up # because she's kind of a uhm@0 stressacoglioni@it .  
513 \*KEL: in the sense that she's really difficult to:  
514 [>] deal with (be)cause she's really dramatic .  
515 \*EMM: mmhm@0 [<] .  
516 \*KEL: but <I d(o)> [//] I will say that her complaints  
517 are quite valid # even when she comes in here  
518 and bust [//] she used to bust my balls # over  
519 the name thing .  
520 \*KEL: she's the one who was a <man for months> [>] .  
521 \*EMM: <oh@0 yes> [<] <yes &=laughs> [>] .  
522 \*KEL: &=laughs [<] .  
523 \*KEL: it's really funny # uhm@0 but [>] +...  
524 \*EMM: Stephen@0 [<] Quinn@0 .  
525 \*KEL: yes # it was really really funny .  
526 \*KEL: anyhow [>] # she really is a &=blows stressapalle@it .  
527 \*EMM: mmhm@0 [<] .  
528 \*KEL: <but she:> [///] # I will say that she's got valid  
529 complaints because it's really not acceptable that  
530 her name was wrong in the system <for so long>  
531 [//] <that it took so long to> [>] fix .  
532 \*EMM: <oh@0 yes # yes> [<] .  
533 \*KEL: <it's one> [>] thing if you show up because nobody  
534 can really know if you're a man or a woman  
535 just from your name if it makes sense .  
536 \*EMM: true [<] .  
537 \*KEL: and that's ok@0 .  
538 \*KEL: and she wasn't upset that it was wrong .  
539 \*KEL: it [///] she was upset that it took like two and  
540 a half months to correct .  
541 \*KEL: &e: <that's the kind of thing that> [>] +...  
542 \*EMM: &=sighs yeah [<] .  
543 \*KEL: it's justifiable .  
544 \*KEL: I mean I would be upset too # saying +"/.  
545 \*KEL: +" you know <it's not really &b> [//] it's not  
546 really that big of a deal .  
547 \*KEL: +" but off [///] on my grade transcripts it shows  
548 up that I'm # Frank@0 .  
549 \*KEL: and I'm applying for I don't know PhD or  
550 something or # <you know whatever> [>] .  
551 \*EMM: <that would be a bit scary> [<] .  
552 \*KEL: <so all of it> [>] it's # frustrating .  
553 \*EMM: yes [<] .  
554 \*KEL: so: when she # uhm@0 was upset about [//] when she  
555 got upset with Bianchi@0 I mean I can understand  
556 it .

557 \*KEL: the thing about the absence though is unexcusable  
558 <(be)cause she should> [>] know # that if you have  
559 a game that's ten days long and you're gone for  
560 two # that's +...  
561 \*EMM: mmhm@0 [<].  
562 \*EMM: well <it's five days> [/] # <it's five> [>] days .  
563 \*KEL: <yeah # exactly> [<].  
564 \*EMM: so: <they're gone mmhm@0> [>] .  
565 \*KEL: <but already ten days and> [<] you're gone for two  
566 .  
567 \*EMM: <exactly # sure> [>] .  
568 \*KEL: <and that's twenty per cent> [<] xx .  
569 \*EMM: and they should be able to work that out without  
570 anyone having to remind them .  
571 \*KEL: yeah .  
572 \*EMM: but she was seriously +...  
573 \*KEL: also [>] .  
574 \*KEL: <I think that was> [<] probably the straw that  
575 just broke the camel's back [>] .  
576 \*EMM: actually [<] <because also> [//] because <she'd had>  
577 [//] I think she'd had a message from Lorenza@0  
578 talking about unjustified absences for exams .  
579 \*EMM: and you must bring your certificate .  
580 \*EMM: for some of them like ten days later we still  
581 haven't had a medical certificate <from them> [>] .  
582 \*KEL: mmhm@0 [<].  
583 \*EMM: uhm@0 and then she: [///] uhm@0 # when he said to her  
584 about this twenty per cent thing she let rippance  
585 +"/.  
586 \*EMM: +" oh@0 Italian bureaucracy !  
587 \*EMM: but it's not Italian bureaucracy .  
588 \*EMM: this is just the rule of the programme .  
589 \*KEL: mmhm@0 .  
590 \*EMM: it's not # <a bureaucratic thing> [>] .  
591 \*KEL: <and she knew the signing> [<] up for it .  
592 \*EMM: yeah@0 and surely in Canada@0 they also have similar  
593 rules .  
594 \*EMM: it's not like # <this is particular> [>] .  
595 \*KEL: <I know # I don't know how it> [<] is in Canada@0  
596 .  
597 \*KEL: I know in America@0 that you can't even +...  
598 \*KEL: if you missed your exam you miss your exam .  
599 \*KEL: not <even a doctor's> [>] note .  
600 \*EMM: yes [<].  
601 \*KEL: unless you get in a car accident or are  
602 <delivering a baby> [>] .  
603 \*EMM: sure [<].  
604 \*EMM: yes .  
605 \*KEL: you know unless it's something <really really  
606 serious> [>] # y(ou) [/] you don't +...  
607 \*EMM: <physically # yes> [<].  
608 \*KEL: not just (be)cause you're sick and go <to the  
609 doctor> [>] and get a note .  
610 \*EMM: mmhm@0 [<].  
611 \*KEL: unless you're dying you go to # your exam .  
612 \*KEL: my mum's # friend took an exam with pneumonia .

613 \*EMM: ah@O .  
614 \*KEL: so: # like +...  
615 \*KEL: my mum took the bar exam a day after she'd had  
616 my brother .  
617 \*EMM: yes .  
618 \*KEL: so those <kind of things> [>] .  
619 \*EMM: <yes exactly> [<] .  
620 \*KEL: you know # it just doesn't happen .  
621 \*KEL: so <she could> [/] ## she could shove it .  
622 \*KEL: but really <it is> [/] it is a &l +...  
623 \*KEL: <also the: thing> [/] # also the thing I wonder if  
624 it's on the website .  
625 \*KEL: I don't remember about having fifteen different  
626 teachers # because it really just goes to show  
627 that the programme is really <not very> [//] not  
628 structured very well .  
629 \*EMM: mmhm:@O .  
630 \*KEL: because I don't think that they would complain  
631 about having fifteen different teachers if xx it was  
632 organised properly .  
633 \*EMM: well it seems a shame that # I mean # <they have  
634 all these courses> [>] .  
635 \*KEL: <well that they can't appreciate having it> [<] .  
636 \*EMM: I understand if she was in the class that had  
637 for those operations that they had four different  
638 professors for that one course .  
639 \*EMM: that was a bit excessive .  
640 \*EMM: but then again if you're taking six courses and  
641 they're all taught by two people that's already  
642 twelve .  
643 \*KEL: mmhm@O .  
644 \*EMM: and then you know bring in a couple of guest  
645 speakers .  
646 \*EMM: throw in an extra thing .  
647 \*EMM: <they had> [//] <maybe they had the Battenberg@O  
648 thing> [>] .  
649 \*KEL: <I think that's just the way that it's organised> [<] <(be)cause you  
650 get> [//] (be)cause it's one thing if you get five hundred different  
651 teachers .  
652 \*KEL: but they all have their little: # periods .  
653 \*EMM: mmhm@O [>] .  
654 \*KEL: but [<] then they've got their period .  
655 \*KEL: and <then it> [///] they change teachers .  
656 \*KEL: and then the other teacher comes back .  
657 \*KEL: and that's I think # what gets frustrating <because  
658 the material I'm> [>] sure gets very confusing when  
659 you don't have like a sequence of teachers .  
660 \*EMM: yes exactly .  
661 \*KEL: it's one thing to have guest speakers come in  
662 for a few days and interrupt this ## learning  
663 process .  
664 \*KEL: but I don't [//] I think it's confusing for the  
665 students <to go back and forth between>  
666 [>1] teachers [>2] .  
667 \*EMM: <which reminds me> [<1] +...  
668 \*EMM: you [<2] just reminded me that I still have an

669 outstanding class to book with Grossi@O and Dainese@O  
670 for May .

671 \*EMM: &=sighs .

672 \*KEL: mmhm@O .

673 \*EMM: there's this day that Dainese@O can't do now .

674 \*EMM: and I've completely forgotten about it .

675 \*KEL: last night I was lying in bed .

676 \*KEL: and I just said +"/.

677 \*KEL: +" you know I really don't miss working at Giorgi@O[>] .

678 %com: Kelly names the institution where she used to work. The name has  
679 been replaced with a pseudonym.

680 \*EMM: oh@O [<] I'm sure !

681 \*KEL: because I was thinking about it .

682 \*KEL: and you must have lots of material things going  
683 on right now .

684 \*KEL: I mean some sort +//.

685 \*KEL: there's always some kind of <material issue> [>] .

686 \*EMM: <there's always something> [<] .

687 \*EMM: uhm:@O really at the moment we're focussing on the  
688 concentrations # <because on> [>] # Wednesday he wants  
689 to present # all the concentrations ## to the  
690 students # and get them to decide what they want .

691

692 \*KEL: mmhm@O [<] .

693 \*EMM: no this week for materials hasn't been so bad .

694 \*EMM: it's just been Ruggeri@O and his English lessons .

695 \*EMM: uhm@O # because they haven't needed lots of materials  
696 .

697 \*EMM: but then next week <the: classes begin> [>] .

698 \*KEL: <he should give you a nice present> [<] going away  
699 .

700 \*EMM: uh@O <I doubt it very much> [>] .

701 \*KEL: <like at the end of the term> [<] .

702 \*EMM: mmhm@O .

703 \*EMM: why <for saving face> [>] ?

704 \*KEL: <a &=chuckles> [<] [//] a bottle of whiskey # <a good  
705 one> [>] .

706 \*EMM: <because it would have been> [<] [//] would have been  
707 a bruttissima@it figura@it if he'd given those things  
708 out as he'd had them because they were really ###  
709 pants useless .

710 \*EMM: anyway .

711 \*KEL: &=laughs .

712 \*EMM: oh@O well what else has been happening ?

713 \*EMM: what about skiing ?

714 \*KEL: oh@O it sucks .

715 \*KEL: it's a big [>] # slushie .

716 \*EMM: why ?

717 \*EMM: was it because it was not enough snow ?

718 \*KEL: yeah well # it's warm now .

719 \*KEL: it's consistently sunny [>] .

720 \*EMM: it's [<] melty .

721 \*KEL: yeah it's like skiing in a granita@it .

722 \*EMM: mmhm@O .

723 \*KEL: and: it's actually dangerous to do that because  
724 you have <so much> [>1] tension on your knees #



725 and: [>2] # you can't turn # in slush .  
726 \*EMM: <oh@0 yes> [<1] .  
727 \*EMM: mmhm@0 [<2] .  
728 \*KEL: and so you can <really hurt yourself> [>] .  
729 \*EMM: mmhm@0 .  
730 \*KEL: so we: +...  
731 \*KEL: it was horrible because probably we threw away  
732 about seventy euros .  
733 \*EMM: mmhm@0 [>] .  
734 \*KEL: <I mean> [<] # I don't really care because we  
735 don't go that often .  
736 \*KEL: so it's not like # I'm gonna [:going to] cry over  
737 it .  
738 \*KEL: but # it was a pain because we &ew[/] went and we  
739 never get out there really early because # we  
740 always go # on Saturday .  
741 \*KEL: and after having worked really hard all week [>1] #  
742 Saturday I've got a hard time getting up at six  
743 o'clock in the morning [>2] .  
744 \*EMM: mmhm@0 [<1] .  
745 \*EMM: understandably [<2] !  
746 \*KEL: yeah .  
747 \*KEL: so we always get out there at about nine thirty  
748 .  
749 \*KEL: and so it means that we're +//.  
750 \*KEL: if we're get there at nine thirty by the time  
751 we go up the hill on the: # gondola .  
752 \*KEL: I don't know what they call it in Italian .  
753 \*KEL: but in American English we call it +...  
754 \*KEL: that little box thing that goes up the hill on  
755 the: wire # <we call it> [>] a gondola .  
756 \*EMM: yeah [<] .  
757 \*EMM: oh@0 !  
758 \*KEL: and I don't know exactly what it's called .  
759 \*KEL: I don't remember in +...  
760 \*EMM: I think <it's called> [//] # <in English it's> [///]  
761 (be)cause I mean like a ski car .  
762 \*KEL: and then we've got the: uhm@0 [///] the &ew [///] <a  
763 ski car> [<] ?  
764 \*EMM: <what's it called that thing> [<] .  
765 \*EMM: no what's it called ?  
766 \*EMM: the # ski lift .  
767 \*KEL: yeah but that's the one +...  
768 \*KEL: we have the ski lift too .  
769 \*KEL: but that's not the one that's closed like a box  
770 .  
771 \*KEL: that's the one that you sit .  
772 \*KEL: that's like a <bench for us> [>] .  
773 \*EMM: oh:@0 [<] !  
774 \*KEL: which is strange .  
775 \*KEL: and then we also have the other one that's much  
776 larger that fits like I don't know fifteen or  
777 twenty people .  
778 \*KEL: maybe even more it's huge .  
779 \*KEL: <that has a> [>] completely different name .  
780 \*EMM: mmhm@0 [<] .

781 \*KEL: the gondola is the one that fits like ## eight  
782 people .  
783 \*KEL: and it's <a little box> [//] # a little square box  
784 .  
785 \*EMM: yes [>] .  
786 \*KEL: <and it's got> [<] like a little # bench that goes  
787 all the way around it and a pole in the middle  
788 .  
789 \*KEL: and you sit there .  
790 \*KEL: and you go all the way up the hill .  
791 \*KEL: and it's like a twenty minute ride [>] .  
792 \*EMM: mmhm:@0 [<] .  
793 \*EMM: not a funicular ?  
794 \*KEL: yeah # ok@0 # it's one of those .  
795 \*EMM: the <funiculi@it funiculà@it> [>] .  
796 \*KEL: <it's like that> [<] # yeah .  
797 \*KEL: &=sings [>] .  
798 %com: Kelly hums a popular Neapolitan song about a funicular.  
799 \*EMM: <yeah # yes> [<] .  
800 \*KEL: uhm:@0 .  
801 \*EMM: there's one uhm:@0 # on: # Lago@it di@it Como@0 .  
802 \*KEL: there's also one uhm@0 in Bergamo@0 # <to go to the:  
803 [/] the historic> [>] centre up the hill .  
804 \*EMM: <uhm@0 yes # that's right> [<] .  
805 \*EMM: yeah .  
806 \*KEL: as <the first> [>] funicular: [/] funic(ular) +...  
807 \*EMM: yeah [<] .  
808 \*KEL: no I meant in # on Montmartre@0 # <in: French [/] in  
809 French> [/] [>1] <in France> [>2] &=laughs .  
810 \*EMM: <uhm@0 yes there's one there> [<1] .  
811 \*EMM: <in France> [<2] .  
812 \*EMM: yes I've been up that one too [>] .  
813 \*KEL: yeah [<] .  
814 \*KEL: I love Montmartre@0 [>] .  
815 \*EMM: <there we are> [<] .  
816 \*EMM: they are good !  
817 \*EMM: <yes it's nice # very nice> [>] .  
818 \*KEL: <yeah # it's beautiful # beautiful> [<] .  
819 \*KEL: I like the little painters' area where they've  
820 <got the uhm: Dali@0 museum> [>] .  
821 \*EMM: <mmhm:@0 # yes yeah> [<] .  
822 \*KEL: yeah I like that .  
823 \*EMM: ah@0 nice .  
824 \*KEL: uhm@0 .  
825 \*KEL: anyway so skiing sucked .  
826 \*KEL: but it was alright .  
827 \*KEL: oh:@0 !  
828 \*EMM: how is Fabrizio's@02 granny ?  
829 \*KEL: <oh@0 she's wonderful> [=! whispers] .  
830 \*KEL: she's so: funny .  
831 \*KEL: <she really is getting> [//] she's turning into a  
832 crochety old lady because we went +...  
833 \*KEL: we stopped on Friday in Trento@0 # to have dinner  
834 with her and the badante@it # serves the guests first .  
835  
836 \*EMM: was she made up ?

837 \*KEL: mmhm@O uh uhm@O kind of .  
838 \*KEL: she's on a diet .  
839 \*EMM: uh@O oh@O !  
840 \*KEL: and # I mean she's lost about # six or seven  
841 kilos or something which I noticed when she went  
842 to the we went to the wedding last [///] # or  
843 two weekends ago .  
844 \*EMM: oh@O .  
845 \*KEL: Fabrizio's@O2 cousin got married [>] .  
846 \*EMM: mmhm@O [<] .  
847 \*KEL: anyway # and I noticed that she lost weight .  
848 \*KEL: and she looked really nice for # this lady's #  
849 normal .  
850 \*KEL: I mean she's not an ugly lady .  
851 \*KEL: but # she looks very ## uhm@O Moldavian .  
852 \*KEL: and she's very Eastern European looking .  
853 \*KEL: anyhow # uhm@O she lost some weight .  
854 \*KEL: and she looked nice .  
855 \*KEL: and then so ## last week I noticed that she was  
856 on a diet (be)cause she didn't eat any pasta@O .  
857 \*KEL: she just had salad and <some vegetables> [>] # no  
858 anything .  
859 \*EMM: ah@O [<] .  
860 \*KEL: and I don't even think she used much oil just  
861 a bit anyway .  
862 \*KEL: uhm@O but nonna@it is really crochety because of  
863 course she's going to serve the guest first .  
864 \*KEL: she only has two hands .  
865 \*EMM: mmhm:@O [>] .  
866 \*KEL: <so she comes> [<] .  
867 \*KEL: and she brings Fabrizio@O and I our plates of  
868 pasta@O .  
869 \*KEL: and then nonna@it goes +"/.  
870 \*KEL: +" oh you're going to serve me too right ?  
871 \*EMM: aw:@O [>] .  
872 \*KEL: <there's some pasta@O> [<] for me too [>] ?  
873 \*EMM: aw:@O [<] .  
874 \*KEL: and so now this crochety old woman now not only  
875 does she want to give this poor: badante@it a hard  
876 time but she also wants to be served first all  
877 the time .  
878 \*KEL: it's <really really funny> [>] .  
879 \*EMM: amazing [<] .  
880 \*EMM: mmhm@O [>] .  
881 \*KEL: she's # completely [<] done a one eighty .  
882 \*KEL: and like she doesn't want to: +...  
883 \*KEL: we've played cards .

## References

- ADALAR, N., TAGLIAMONTE, S. (1998) Borrowed nouns, bilingual people: the case of “Londrali” in Northern Cyprus. *International Journal of Bilingualism*, 2:2:139-160.
- AITCHISON, J. (2003) *Words in the Mind: An introduction to the Mental Lexicon* (Third edition). Oxford: Blackwell Publishing.
- ALFONZETTI, G. (1998) The conversational dimension in code-switching between Italian and dialect in Sicily. In P. Auer (ed) *Code-Switching in Conversation: Language, interaction and identity*, pp. 180-211. London: Routledge.
- ÁLVAREZ-CÁCCAMO, C. (1998) From ‘switching code’ to ‘code-switching’: towards a reconceptualization of communicative codes. In P. Auer (ed) *Code-Switching in Conversation: Language, interaction and identity*, pp. 29-48. London: Routledge.
- ANTONELLI, G. (2007) *L'italiano nella società della comunicazione*. Bologna: il Mulino.
- AUER, P. (1984) *Bilingual Conversation*. Amsterdam: Benjamins.
- AUER, P. (1995) The pragmatics of code-switching: a sequential approach. In L. Milroy, P. Muysken (eds) *One speaker, two languages: Cross-disciplinary perspectives on code-switching*, pp. 115-135. Cambridge: Cambridge University Press.
- AUER, P. (ed) (1998) *Code-switching in conversation: Language, interaction and identity*. London: Routledge.
- AUER, P. (1999) From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism*, 3:4:309-332.
- AUER, P. (2000) Why should we and how can we determine the “base language” of a bilingual conversation? *Estudios de Sociolingüística*, 1:1:129-144.
- BACKUS, A. (1996) Two in one: *Bilingual speech of Turkish immigrants in The Netherlands*. Tilburg: Tilburg University Press.
- BACKUS, A. (2003) Units in code switching: evidence for multimorphemic elements in the lexicon. *Linguistics*, 41:1:83-132.
- BACKUS, A. (2005) Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9:3/4:307-340
- BACKUS, A., DORLEIJN, M. (2009) Loan translation versus code-switching. In B.E. Bullock, A.J. Toribio (eds) *The Cambridge Handbook of Linguistic Code-switching*, pp. 75-93. Cambridge: Cambridge University Press.
- BARNETT, R., CODÓ, E., EPPLER, E., FORCADELL, M., GARDNER-CHLOROS, P., VAN HOUT, R., MOYER, M., TORRAS, M. C., TURELL, M. T. SEBBA, M., STARREN, M., WENSING, S. (2000) The LIDES Coding Manual: A document for preparing and analyzing language interaction data Version 1.1—July, 1999. *International Journal of Bilingualism*, 2000:4:2.
- BELAZI, H., RUBIN, E., TORIBIO, A. J. (1994) Code-switching and X-bar theory: the Functional Head Constraint. *Linguistic Inquiry*, 25:221-238.
- BENTAHILA, A., DAVIES, E. (1995) Patterns of code-switching and patterns of language contact. *Lingua*, 96:75-93.

- BENTLEY, D. (2002) New Linguistic Research into Italo-Romance Dialects. In A. L. Lepschy, A. Tosi (eds) *Multilingualism in Italy: Past and Present*, pp. 82-107. Oxford: Legenda.
- BERTINETTO, P. M., LOPORCARO, M. (2005) The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, 35:2:131-151.
- BETTONI, C. (2010) Tra lingua, dialetto e inglese: mezzo secolo di emigrazione italiana in Australia. In A. Ledgeway, A.L. Lepschy (eds) *Into and out of Italy: lingua e cultura della migrazione italiana*, pp. 153-163. Perugia: Guerra Edizioni.
- BETTONI, C., RUBINO, A. (1996) *Emigrazione e comportamento linguistico: un'indagine sul trilinguismo dei siciliani e dei veneti in Australia*. Lecce: Congedo Editore.
- BIALYSTOK, E., CRAIK, F. I. M., GREEN, D. W., GOLLAN, H. T. (2009) Bilingual minds. *Psychological Science in the Public Interest*, 10:3:89-129.
- BIZZONI, F. (2003) Il "Matrix Language Frame Model": un'applicazione all'italiano di immigranti di 1a generazione in Messico. In A. De Fina, F. Bizzoni (eds) *Italiano e italiani fuori d'Italia*, pp. 69-94. Perugia: Guerra Edizioni.
- BLOM, J. P., GUMPERZ, J. J. (1972) Social meaning in linguistic structure: code-switching in Norway. In J. J. Gumperz, D. Hymes (eds) *Directions in Sociolinguistics: The Ethnography of Communication*, pp. 407-434. New York: Holt, Rinehart and Winston.
- BRÖERSMA, M. (2009) Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*, 12:4:447-462
- BUDZHAK-JONES, S. (1998) Against word-internal codeswitching: Evidence from Ukrainian-English bilingualism. *International Journal of Bilingualism*, 2:2:161-182.
- BULLOCK, B. E., TORIBIO, A. J. (eds) (2009) *The Cambridge Handbook of Linguistic Code-switching*. Cambridge: Cambridge University Press.
- CANDEA, M., VASILESCU, I., ADDA-DECKER, M. (2005) Inter- and intra-language acoustic analysis of autonomous fillers. Proceedings of DISS 05, *Disfluency in Spontaneous Speech Workshop*, 10-12 September, Aix-en-Provence.
- CANTONE, K. F. (2007) *Code-switching in Bilingual Children*. Studies in Theoretical Psycholinguistics, 37. Dordrecht: Springer.
- CERVI, B. (1991) The Italian speech community. In S. Alladina, V. Edwards (eds) *Multilingualism in the British Isles: the older mother tongues and Europe*, pp. 216-228. London: Longman.
- CHAN, B. H. (2003) *Aspects of the Syntax, the pragmatics and the production of code-switching: Cantonese and English*. New York: Peter Lang.
- CHAN, B. H. (2009) Code-switching between typologically distinct languages. In B.E. Bullock, A.J. Toribio (eds) *The Cambridge Handbook of Linguistic Code-switching*, pp. 182-198. Cambridge: Cambridge University Press.
- CHESHIRE, J., GARDNER-CHLOROS, P. (1998) Code-switching and the Sociolinguistic Gender Pattern. *International Journal of the Sociology of Language, Special edition on Women's Languages in Various Parts of the World*, 129:5-34.
- CLARK, H., FOX TREE, J.E. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition*, 84:1:73-111.
- CLYNE, M. (1987) Constraints on code-switching: how universal are they? *Linguistics*, 25:739-764.

- CLYNE, M. (2003) *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.
- COLUZZI, P. (2004) Regional and Minority Languages in Italy. A general introduction on the present situation and a comparison of two case studies: language planning for Milanese (Western Lombard) and Friulian. *Working Papers 14. Barcelona: Mercator and Escarré International Centre for Ethnic Minorities and Nations*.
- CORLEY, M., STEWART, O. (2008) Hesitation Disfluencies in Spontaneous Speech: The Meaning of *um*. *Language and Linguistics Compass*, 2:4:589-602.
- COSTA, A., SANTESTEBAN, M. (2004) Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50:491-511.
- CRYSTAL, D. (2003) *English as a Global Language. Second edition*. Cambridge: Cambridge University Press.
- DARDANO, M. (1986) The influence of English on Italian. In W. Viereck, B. Wolf-Dietrich (eds) *English in Contact with other languages*, pp. 231-252. Budapest: Akadémiai Kiadó.
- DE FINA, A. (2007) Code-switching and the construction of ethnic identity in a community of practice. *Language in Society*, 36:3:371-392.
- DE LEEUW, E. (2007) Hesitation Markers in English, German, and Dutch. *Journal of Germanic Linguistics*, 19:2:85-114.
- DE MAURO, T. (1993) (ed) *Lessico di frequenza dell'italiano parlato*. Milano: Etas Libri.
- DEUCHAR, M. (2005) Congruence and Welsh-English code-switching. *Bilingualism: Language and Cognition*, 8:3:255-269.
- DEUCHAR, M. (2006) Welsh-English code-switching and the Matrix Language Frame model. *Lingua*, 116:11:1745-2022.
- DEUCHAR, M. (in press) Code-switching. In C. A. Chapelle (ed) *Encyclopaedia of Applied Linguistics*. Oxford: Blackwell Publishing.
- DEUCHAR, M., DAVIES, P. (2009) Code switching and the future of the Welsh language. *International Journal of the Sociology of Language*, 195:15-38.
- DEUCHAR, M., DAVIES, P., HERRING, J., PARAFITA COUTO, M. C., CARTER, D. (forthcoming) Code-switching corpora: A state of the art. In E. Thomas, I. Mennen (eds) *Unraveling Bilingualism: A Cross-Disciplinary Perspective*. Bristol: Multilingual Matters.
- DEUCHAR, M., MUYSKEN, P., WANG, S. (2007) Structured Variation in Codeswitching: Towards an Empirically Based Typology of Bilingual Speech Patterns. *International Journal of Bilingual Education and Bilingualism*, 10:3:298:340.
- DEWAELE, J-M. (2001). Activation or inhibition? The interaction of L1, L2 and L3 on the language mode continuum. In U. Jessner, B. Hufeisen, X. Cenoz (eds) *Cross-linguistic Influence in Third Language Acquisition: psycholinguistic perspectives*, pp. 69-89. Clevedon: Multilingual Matters.
- DI SCIULLO, A.M., MUYSKEN, P., SINGH, R. (1986) Government and code-mixing. *Journal of Linguistics*, 22:1-24.
- EDWARDS, D. (2006) Deviant case analysis. In V. Jupp (ed) *The Sage Dictionary of Social Research Methods*, pp. 68-69. London: Sage Publications.

- EPPLER, E. (1994) Code-switching in 'Emigranto'. *VIEWS - Vienna English Working Papers*, 3:2:75-93.
- EXTRA, G., YAĞMUR, K. (2004) *Urban Multilingualism in Europe: Immigrant Minority Languages at Home and School*. Clevedon: Multilingual Matters.
- EZE, E. (1998) Lending credence to a borrowing analysis: Lone English-origin incorporations in Igbo discourse. *International Journal of Bilingualism*, 2:2:183-202.
- FEHRINGER, C., FRY, C. (2007) Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica*, 41:1-2:37-72.
- FINLAYSON, R., CALTEAUX, K., MYERS-SCOTTON, C. (1998) Orderly mixing and accommodation in South African codeswitching. *Journal of Sociolinguistics*, 2:3:395-420.
- FISHMAN, J. (1965 [2000]) Who speaks what language to whom and when? In Li Wei (ed) *The Bilingualism Reader*, pp. 89-106. London: Routledge.
- FISHMAN, J. (1967 [2000]) Bilingualism with and without diglossia; diglossia with and without bilingualism. In Li Wei (ed) *The Bilingualism Reader*, pp. 81-88. London: Routledge.
- FRANCESCHINI, R. (1998) Code-switching and the notion of code in linguistics: proposal for a dual focus model. In P. Auer (ed) *Code-Switching in Conversation: Language, interaction and identity*, pp. 51-72. London: Routledge.
- FOX, B. A., HAYASHI, M., JASPERSON, R. (1996) Resources and repair: a cross-linguistic study of syntax and repair. In E. Ochs, E. A. Schegloff, S. A. Thompson (ed) *Interaction and grammar*, pp. 185-237. Cambridge: Cambridge University Press.
- FOX, B. A., MASCHLER, Y., UHMANN, S. (2010) A cross-linguistic study of self-repair: Evidence from English, German, and Hebrew. *Journal of Pragmatics*, 42:9:2487-2505.
- GAFARANGA, J. (2000) Medium repair vs. other-language repair: Telling the medium of a bilingual conversation. *International Journal of Bilingualism*, 4:3:327-350.
- GAFARANGA, J. (2001) Linguistic identities in talk-in-interaction: Order in bilingual conversation. *Journal of Pragmatics*, 33:1901-1925.
- GAFARANGA, J. (2007) *Talk in Two Languages*. Basingstoke: Palgrave Macmillan.
- GAFARANGA, J. (2009) The Conversation Analytic model of code-switching. In B. E. Bullock, A. J. Toribio (eds) *The Cambridge Handbook of Linguistic Code-switching*, pp. 114-126. Cambridge: Cambridge University Press.
- GAFARANGA, J., TORRAS, M. (2001) Language versus Medium in the study of bilingual conversation. *International Journal of Bilingualism*, 5:2:195-219.
- GAFARANGA, J., TORRAS, M. (2002) Interactional otherness : Towards a redefinition of codeswitching. *International Journal of Bilingualism*, 6:1:1-22.
- GAMBINO, M. (2010) *Italian evaluative morphology: a syntactic approach*. PhD dissertation, Università degli Studi di Padova (Italy).
- GARDNER-CHLOROS, P. (1995) Code-switching in community, regional and national repertoires: the myth of the discreteness of linguistic systems. In L. Milroy, P. Muysken (eds) *One speaker, two languages: Cross-disciplinary perspectives on code-switching*, pp. 68-89. Cambridge: Cambridge University Press.
- GARDNER-CHLOROS, P. (2007) European immigrant languages. In D. Britain (ed) *Language in the British Isles*, pp. 325-340. Cambridge: Cambridge University Press.

- GARDNER-CHLOROS, P. (2009) *Code-switching*. Cambridge: Cambridge University Press.
- GARDNER-CHLOROS, P., EDWARDS, M. (2004) Assumptions behind grammatical approaches to code-switching: when the blueprint is a red herring. *Transaction of the Philological Society*, 102:1:103-129.
- GARDNER-CHLOROS, P., REEVA, C. CHESHIRE, J. (2000) Parallel patterns? A comparison of monolingual speech and bilingual codeswitching discourse. *Journal of Pragmatics*, 32:1305-1341.
- GARFINKEL, H. (1967) *Studies in Ethnomethodology*. Cambridge: Polity Press.
- GHAFFAR SAMAR, R., MEECHAN, M. (1998) The Null Theory of codeswitching versus the Nonce Borrowing Hypothesis; Testing the fit in Persian-English bilingual discourse. *International Journal of Bilingualism*, 2:2:203-220.
- GIAMPAPA, F. (2001) Hyphenated identities: Italian-Canadian youth and the negotiation of ethnic identities in Toronto. *International Journal of Bilingualism*, 5:3:279-315.
- GOGLIA, F. (2010) Multilingualism in the immigrant context: the case of the Igbo-Nigerians in Padua. In A. Ledgeway, A.L. Lepschy (eds) *Into and out of Italy: lingua e cultura della migrazione italiana*, pp. 55-64. Perugia: Guerra Edizioni.
- GREER, T. (2008) Accomplishing difference in bilingual interaction: Translation as backwards-oriented medium repair. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 27:1-2:99-127.
- GROSJEAN, F. (2008) *Studying bilinguals*. Oxford: Oxford University Press.
- GULLBERG, M., INDEFREY, P., MUYSKEN, P. (2009) Code-switching in Papaminto-Dutch sentence production: evidence from a sentence completion task. Paper presented at the *International Symposium on Bilingualism 7*, Utrecht.
- GUMPERZ, J.J. (1982) *Discourse Strategies*. Cambridge: Cambridge University Press.
- HALMARI, H. (1993) Structural relations and Finnish-English bilingual code-switching. *Linguistics*, 31:6:1043-1068.
- HAUGEN, E. (1950) The analysis of linguistic borrowings. *Language*, 26:210-231.
- HERITAGE, J. (1994) Explanations as accounts: a conversation analytic perspective. In C. Antaki (ed) *Analysing Everyday Explanation: a casebook of methods*, pp. 127-144. London: Sage Publications.
- HLAVAC, J. (2006) Bilingual discourse markers: evidence from Croatian-English code-switching. *Journal of Pragmatics*, 38:1870-1900.
- HLAVAC, J. (2011) Hesitation and monitoring phenomena in bilingual speech: A consequence of code-switching or a strategy to facilitate its incorporation? *Journal of Pragmatics*, 43:3793-3806.
- JONES, M. (2005) Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French. *French Language Studies*, 15:1-23.
- JOSHI, A. (1985) Processing of Sentences with Intrasentential Switching, In: D.R. Dowty, L. Karttunen, A. Zwicky (eds) *Natural Language Parsing*, pp. 190-205. Cambridge: Cambridge University Press.
- KINDER, J.J. (1988) Transference markers in New Zealand Italian: constructing context and negotiating identity. *Australian Review of Applied Linguistics*, 11:1:9-21. (Accessed original manuscript).



- KJELLMER, G. (2003) Hesitation. In Defence of ER and ERM. *English Studies*, 84:2:170-198.
- LABOV, W. (1971) The notion of 'system' in creole languages. In D. Hymes (ed) *Pidginization and Creolization of Languages*, pp. 447-72. Cambridge: Cambridge University Press.
- LABOV, W. (1972) *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- LEPSCHY, A. L., LEPSCHY, G. (1977) *The Italian language today*. London: Hutchinson (2nd edition – London: Routledge, 1988).
- LEPSCHY, G. (2002) What is the Standard? In A. L. Lepschy, A. Tosi (eds) *Multilingualism in Italy: Past and Present*, pp. 74-81. Oxford: Legenda.
- LI WEI (1994) *Three Generations Two Language One Family: Language choice and language shift in a Chinese community in Britain*. Clevedon: Multilingual Matters.
- LI WEI (1998) The 'why' and 'how' questions in the analysis of conversational code-switching. In P. Auer (ed.) *Code-switching in Conversation: Language, Interaction and Identity*, pp. 156-176. London: Routledge.
- LI WEI (2002) 'What do you want me to say?' On the Conversation Analysis approach to bilingual interaction. *Language in Society*, 31:159-180.
- LIDDICOAT, A. J. (2007) *An introduction to Conversation Analysis*. London: Continuum.
- LÜDI, G. (2003) Code-switching and unbalanced bilingualism. In J-M. Dewaele, A. Housen, L. Wei (eds) *Bilingualism: Beyond Basic Principles*, pp. 174-178. Clevedon: Multilingual Matters.
- MACLAY, H., OSGOOD, C. E. (1959) Hesitation phenomena in spontaneous English speech. *Word*, 15:19-44.
- MACSWAN, J. (2000) The architecture of the bilingual language faculty: evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, 3:1:37-54.
- MACSWAN, J. (2004) Code Switching and Grammatical Theory. In T. K. Bhatia, W. C. Ritchie (eds) *The Handbook of Bilingualism*, pp. 283-311. Oxford: Blackwell Publishing.
- MACSWAN, J. (2009) Generative approaches to code-switching. In B.E. Bullock, A. J. Toribio (eds) *The Cambridge Handbook of Linguistic Code-switching*, pp. 309-335. Cambridge: Cambridge University Press.
- MACWHINNEY, B. (2000) *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MAIDEN, M. (1995) *A Linguistic History of Italian*. Harlow: Longman.
- MAIDEN, M. (2002) The definition of multilingualism in historical perspective. In A. L. Lepschy, A. Tosi (eds) *Multilingualism in Italy: Past and Present*, pp. 31-46. Oxford: Legenda.
- MAHOOTIAN, S. (1996) Code-switching and universal constraints: evidence from Farsi/English. *World Englishes*, 15:3:377-384.
- MATRAS, Y. (2009) *Language Contact*. Cambridge: Cambridge University Press.
- MEEUWIS, M. AND BLOMMAERT, J. (1998) A monolectal view of code-switching: Layered code-switching among Zairians in Belgium. In P. Auer (ed) *Code-Switching in Conversation: Language, interaction and identity*, pp. 76-100. London: Routledge.
- MILROY, L. (1987) *Language and social network*. Oxford: Basil Blackwell.

- MILROY, L. (1989) *Observing and analysing natural language*. Oxford: Basil Blackwell.
- MUYSKEN, P. (2000) *Bilingual speech: a typology of code-mixing*. Cambridge: Cambridge University Press.
- MYERS-SCOTTON, C. (1993a) *Duelling Languages: grammatical structures in codeswitching*. Oxford: Clarendon Press.
- MYERS-SCOTTON, C. (1993b) *Social Motivations For Codeswitching: Evidence From Africa*. Oxford: Clarendon Press.
- MYERS-SCOTTON, C. (2002) *Contact linguistics: bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- O'CONNELL, D., KOWAL, S. (2005) Uh and um revisited: are they interjections for signaling delay? *Journal of Psycholinguistic Research*, 34:6:555-76.
- OCHS, E. (1979) Transcription as Theory. In E. Ochs, B. Schieffelin (eds) *Developmental pragmatics*, pp. 43-72. New York: Academic Press.
- OESCH-SERRA, C. (1998) Discourse connectives in bilingual conversation: The case of an emerging Italian-French mixed code. In P. Auer (ed) *Code-Switching in Conversation: Language, interaction and identity*, pp. 101-122. London: Routledge.
- OLSON, D. (2011) Modulating language switching costs: The effect of linguistic context on production of language switches. Paper presented at the *International Symposium on Bilingualism 8*, Oslo.
- OWENS, J. (2005) Hierarchicalised matrices: codeswitching among urban Nigerian Arabs. *Linguistics*, 43:5:957-993.
- PANESE, M. (1992) Il Code-switching come strategia comunicativa: un'indagine nella comunità italiana a Londra. In A. A. Sobrero (ed) *Il dialetto nella conversazione: ricerche di dialettologia pragmatica*, pp. 43-78. Lecce: Congedo Editore.
- PASQUANDREA, S. (2007) *Code-switching e identità: pratiche discorsive di famiglie italiane in paesi anglofoni*. PhD dissertation, Università degli Studi di Pisa.
- PIETROPAOLO, D. (2010) Language loyalty and the culture of immigration in early Canadian Italian. In A. Ledgeway, A. L. Lepschy (eds) *Into and out of Italy: lingua e cultura della migrazione italiana*, pp. 119-128. Perugia: Guerra Edizioni.
- PIRKKALAINEN, S. (2002) Perché business is business: A study of Anglicisms in *Milano Finanza*. MA Dissertation, University of Jyväskylä (Finland).
- POPLACK, S. (1980, [2000]) Sometimes I'll start a sentence in Spanish *y termino en Español*: toward a typology of code-switching. In Li Wei (ed) *The Bilingualism Reader*, pp. 221-256. London: Routledge.
- POPLACK, S. (1982) Bilingualism and the vernacular. In A. Valdman, B. Hartford (eds) *Issues in International Bilingual Education: The Role of the Vernacular*, pp. 1-24. New York: Plenum Publishing Co.
- POPLACK, S. (1985) Contrasting patterns of code-switching in two communities. *Papers from the 5th International Conference on Methods in Dialectology*. University of Victoria (BC). Accessed at <http://hdl.handle.net/10315/6618>.

- POPLACK, S. (1988) Contrasting patterns of code-switching in two communities. In M. Heller (ed) *Codeswitching. Anthropological and Sociolinguistic Perspectives*, pp. 215-224. Berlin: Mouton de Gruyter.
- POPLACK, S. (1993) Variation Theory and language contact. In D. Preston (ed) *American dialect research: an anthology celebrating the 100th anniversary of the American Dialect Society*, pp. 251-286. Amsterdam: Benjamins.
- POPLACK, S. (2004) Code-switching. In U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (eds) *Soziolinguistik. An international handbook of the science of language*, 2nd edition, pp. 28-48. Berlin: Walter de Gruyter.
- POPLACK, S., MEECHAN, M. (1998) How languages fit together in codemixing. *International Journal of Bilingualism*, 2:2:127-138.
- POPLACK, S. WHEELER, S., WESTWOOD, A. (1989) Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. *World Englishes*, 8:3:389-406.
- POULISSE, N., BONGAERTS, T. (1994) First language use in second language production. *Applied Linguistics*, 15:36-57.
- RADATZ, H.-I., STAMMERJOHANN, H. (1996) De la rhétorique à la typologie et vice versa: le jeu de la position de l'adjectif dans les langues romanes. *Studi italiani di linguistica teorica e applicata*, 25:3:521-531.
- RIEGER, C. L. (2003) Repetitions as self-repair strategies in English and German conversations. *Journal of Pragmatics*, 35:1:47-69.
- ROCCHI, L. (2008) «...Eravamo anche cinque o sei, tutte fimmine... e 'ndammo up to London e... and that was the first time I've been in London...». Reti sociali e scelte linguistiche di emigrati italiani in ambiente anglofono. *Studi Linguistici e Filologici Online*, 6:219-273.
- ROMAINE, S. (1995) *Bilingualism. Second Edition*. Oxford: Blackwell Publishing.
- RUBINO, A. (2003) Trilinguismo e monolinguisimo tra gli italo-australiani: due famiglie a confronto. In A. De Fina, F. Bizzoni (eds) *Italiano e italiani fuori d'Italia*, pp. 145-174. Perugia: Guerra Edizioni.
- SACKS, H., SCHEGLOFF, E., JEFFERSON, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:4:696-735.
- SANKOFF, G. (2001) Linguistic Outcomes of Language Contact. In P. Trudgill, J. Chambers, N. Schilling-Estes (eds) *Handbook of Sociolinguistics*, pp. 638-668 Oxford: Basil Blackwell.
- SANKOFF D., POPLACK, S. (1981) A formal grammar for code-switching. *Papers in Linguistics*, 14:3-46.
- SAVILLE-TROIKE, M. (2003) *The Ethnography of Communication: An introduction* (Third Edition). Oxford: Blackwell Publishing.
- SCAGLIONE, S. (2003) Segnali discorsivi allogeni nelle varietà di emigrazione: you know, and, so, well nell'italiano di San Francisco. In A. De Fina, F. Bizzoni (eds) *Italiano e italiani fuori d'Italia*, pp. 45-68. Perugia: Guerra Edizioni.
- SCHEGLOFF, E. (2010) Some Other "Uh(m)"s. *Discourse Processes*, 47:2:130-174.
- SCHEGLOFF, E., JEFFERSON, G., SACKS, H. (1977) The preference for self-correction in the organization of repair in conversation. *Language*, 53:2:361-382.
- SCHMID, S. (2005) Code-switching and Italian abroad. Reflections on language contact and bilingual mixture. *Rivista di Linguistica*, 17:1:113-165.

- SMITH, D. J. (2006) Thresholds leading to shift: Spanish/English codeswitching and convergence in Georgia, USA. *International Journal of Bilingualism*, 10:2:207-240.
- STAMMERS, J. (2009) *The integration of English-origin verbs in Welsh*. PhD dissertation, Bangor University.
- STELL, G. (2009) Codeswitching and ethnicity: grammatical types of codeswitching in the Afrikaans speech community. *International Journal of the Sociology of Language*, 199:103-128.
- TABOURET KELLER, A. (1995) Conclusion: code-switching research as a theoretical challenge. In L. Milroy, P. Muysken (eds) *One speaker, two languages: Cross-disciplinary perspectives on code-switching*, pp. 68-89. Cambridge: Cambridge University Press.
- TAMBURELLI, M. (2010) The vanishing languages of Italy: Diglossia, bilingualism, and language shift. Keynote address at *Languages in Contact 2010*, Philological School of Higher Education, Wrocław, Poland.
- TORIBIO, A. J. (2004) Spanish/English Speech Practices: Bringing Chaos to Order. *International Journal of Bilingual Education and Bilingualism*, 7:2-3:133-154.
- TOSI, A. (1984) *Immigration and bilingual education*. Oxford: Pergamon Press.
- TOSI, A. (1991) *Italian overseas: The language of Italian communities in the English-speaking world*. Firenze: Giunti.
- TOSI, A. (2001) *Language and society in a changing Italy*. Bristol: Multilingual Matters.
- TOSI, A. (2004) *The language situation in Italy*. *Current Issues in Language Planning*, 5:3:247-335.
- TOSI, A. (2010) The language of Italians in Britain. In A. Ledgeway, A.L. Lepschy (eds) *Into and out of Italy: lingua e cultura della migrazione italiana*, pp. 99-110. Perugia: Guerra Edizioni.
- TREFFERS-DALLER, J. (2009) Code-switching and transfer: an exploration of similarities and differences. In B. E. Bullock, A. J. Toribio (eds) *The Cambridge Handbook of Linguistic Code-switching*, pp. 58-74. Cambridge: Cambridge University Press.
- TURPIN, D. (1998) 'Le Français, c'est le *last frontier*': The Status of English-origin Nouns in Acadian French". *International Journal of Bilingualism*, 2:2:203-219.
- VASILESCU, I., CANDEA, M., ADDA-DECKER, M. (2005) Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages. Proceedings of *Interspeech 2005*, Lisbon.
- VILLATA, B. (2003) L'Italiano in Canada: storia e prospettive. In A. De Fina, F. Bizzoni, (eds) *Italiano e italiani fuori d'Italia*, pp. 177-198. Perugia: Guerra Edizioni.
- WEINREICH, U. (1953) *Languages in contact*. New York: Mouton.
- WINFORD, D. (2003) *An Introduction to Contact Linguistics*. Oxford: Blackwell.

