Multi-word insertions in code-mixed utterances: Italian-German code mixing in South Tyrol

Simone Ciccolone

(University of Cagliari)

Abstract

The present study proposes a corpus-based analysis of the distribution of insertional code mixing (cf. Muysken 2000) occurring in a corpus of bilingual speech collected in South Tyrol. The analysis uses two surface parameters (directionality and extension), not restrained to a specific theoretical model, to classify the switches. The aim of this study is to observe recurring patterns and functions in the combined use of codes within the utterance, in particular by comparing the functions and distribution of single-word switches with those of multi-word insertions (e.g. phrases, word combinations and idioms). The discussion shows the emergence of two macro-functional patterns consistently attested in both types of switches.

Key Words – code mixing; bilingual speech; South Tyrol

Il presente contributo intende proporre un'analisi *corpus-based* dei fenomeni di contatto linguistico nel discorso, con particolare attenzione all'enunciazione mistilingue di tipo insertivo (cfr. Muysken 2000) attestati in un corpus di parlato bilingue raccolto in Alto Adige. L'analisi utilizza due parametri di superficie, indipendenti da specifici modelli teorici relativi al code mixing, per l'individuazione e classificazione dei fenomeni di enunciazione mistilingue: direzionalità ed estensione. Lo scopo dell'analisi è osservare la presenza di schemi combinatori ricorrenti nell'uso combinato dei codici all'interno dell'unità enunciato, in particolare confrontando la distribuzione e le funzioni dei casi di enunciazione mistilingue monològa e dei casi di stile insertivo esteso su più elementi lessicali (tra cui sintagmi, locuzioni e costruzioni idiomatiche).

Parole chiave – enunciazione mistilingue; parlato bilingue; Alto Adige

1. Introduction[1]

Bilingual speech communities are constantly confronted with the evidence of linguistic diversity, which deeply affects their language habits and attitudes. Despite the higher degree of diversity that the co-existence of multiple codes may imply, language contact phenomena in different bilingual communities show interesting similarities. Muysken's model (2000) for the classification of code-switching phenomena constitutes not only a valuable example of how similar patterns might emerge in different contact situations and for different language pairs, but also an insightful account of the complex and multifold nature of this class of phenomena, which is inevitably related to the contact situation and the relationship between the languages (or language varieties) involved.

Generally speaking, prototypical code-switching phenomena tend to assume a well-known set of conversational functions (cf. Auer 1988, 1999), representing an additional conversational strategy available to bilinguals. Nonetheless, when the switch occurs within a single utterance, it seems rather hard to attribute a clear intentional meaning to the switch itself, in particular when there are no further signs of transition that might help the observer to discern between a conversationally meaningful choice and the occasional transfer of isolated items from the lexicon of the other language (cf. Poplack 1987 among others).

The analysis presented here focuses on code mixing in German-Italian bilingual speech in South Tyrol (see Section 3). The quantitative data on the distribution of investigated phenomena as well as the excerpts used for qualitative observation are drawn from the *Kontatto* bilingual spoken corpus[2].

The aim of this study is to observe the distribution of code-mixing phenomena within the corpus and the interaction between languages within the utterance or group of connected utterances (see Section 2). The discussion focuses on multi-word insertions and their function within the utterance, with the goal of highlighting possible patterns in the combined use of the codes at a macro-syntactic and textual level.

Section 2 provides a definition of the specific object and units of analysis considered in this study; Section 3 presents in more detail the observed contact situation and the data; Section 4 discusses examples of code-mixing phenomena from the corpus.

2. Object and unit of analysis

The emergence of recurring patterns in the combined use of languages in bilingual speech will not be discussed from the perspective of its conversational, «locally meaningful» function (cf. Auer 1999: 310), but rather from a more systematic and context-independent point of view concerning the macro-syntactic and textual structure. To this purpose, the term *code mixing* has been chosen as opposed to *code switching*, in agreement with Berruto's suggestion of defining it as «everything in code switching that is relevant to

---

[2] *Kontatto* corpus, by Silvia Dal Negro (Free University of Bozen/Bolzano); research project: "Italiano-tedesco: aree storiche di contatto in Sudtirolo e Trentino" (06.2011-05.2014); cf. Dal Negro and Ciccolone (2018a).

syntactic rules and principles» (Berruto 2011: 48, my translation), but extending its scope to information structure and text-oriented rules.

As Berruto (2009) points out, the choice of the unit *per se* is a key condition to the type of phenomena as well as to the level of analysis that one might observe. In this paper as well as in the data structure of the *Kontatto* corpus, the adopted unit of analysis is defined as follows:

> a contiguous stretch of speech produced by a single speaker without clearly discernible interruptions by the interlocutors, representing at least a single illocutionary act or at the most a sequence of consecutive connected speech acts pertaining to the same conversational move, thus producing a meaningful progression in the interaction.

For reasons of brevity, in the following sections this unit will sometimes be referred to as *utterance*, but the latter will have to be seen as including the combination of consecutive connected speech acts expressed by chunks not recognisable as separate utterances themselves. This working definition has been provisionally adopted with the purpose to address actual occurrences of speech units showing some continuity in form and meaning, at least at a higher intonational and informational level.

Clearly, the chosen unit does not correspond, strictly speaking, to syntactic units such as sentence or clause; therefore, the definition of code mixing used here is not co-extensive with either intrasentential switching (cf. Poplack 1980) or intraclausal switching (cf. Deuchar et al. 2018). Further investigation of the corpus is required to obtain fully comparable figures based on these syntactic measures. A sharper definition of the unit of analysis and the discussion of the theoretical consequences of its choice might be found in Ciccolone and Dal Negro (2021).

3. Data and method

South Tyrol is a multilingual region in Northern Italy with three official languages: Italian, German and Ladin. Ladin (which will not be treated here) is restricted mainly to the valleys of Gardena and Badia; Italian L1 speakers concentrate in the biggest urban centres (e.g. Bolzano), while German L1 speakers are spread through the whole province and are the majority in almost all municipalities[3].

As pointed out by Mioni (1990), there are three different speech communities in the region (not considering Ladin): (a) Italian L1 speakers, mostly monolingual (Standard German is learned at school but scarcely used in everyday communication); (b) speakers of local German varieties as L1 (South Tyrolean German, a subset of Bavarian dialects) and Italian as L2, actively used in intergroup communication; (c) speakers in the Bassa Atesina area[4], traditionally bilingual and with a local Italo-romance variety (Trentino) unattested in the rest of South Tyrol (cf. Dal Negro 2018).

The *Kontatto* corpus involves speakers from (b) and (c) with variable degrees of bilingualism. Due to its peculiar contact situation, a large majority of the recordings and speakers involved come from Bassa Atesina. Data collection and transcription were

---

[3] Data from speech group declarations in 2011 Census; cf. ASTAT (2017).
[4] The area of Bassa Atesina includes the urban centres on the eastern bank of the Adige River south of Bolzano, such as Laives/Leifers, Bronzolo/Branzoll, Salorno/Salurn (included in our field research) as well as Ora/Auer, Magré/Margreid and Termeno/Tramin.

conducted by members of the same speech community and were supervised by the research team (cf. Dal Negro and Ciccolone 2018a for further details).

The corpus consists of 55 audio recordings of interactions among bilinguals, including: (a) spontaneous speech interactions in everyday situations (e.g. family dinner, chat among friends); (b) free informal interviews with collectors' friends or relatives; (c) semi-spontaneous speech elicited through the *Map Task* technique.

The total recording time is approximately 18 hours. The recordings were fully transcribed and manually tagged on a word-by-word level by part of speech and language. The transcription files were created through ELAN[5]; the main transcription line is segmented into units roughly corresponding to single or clusters of utterances (according to the definition given in Section 2). Table 1 shows the statistics of the corpus and the distribution of languages therein.

Table 1. Statistics of the *Kontatto* corpus

Recordings	55	
Total recording time	18:02 h	
Total no. of words	146,719	
- German	117,916	(80.4%)
- Italian	16,864	(11.5%)
- Trentino	3,893	(2.6%)
- interjections/fillers	4,666	(3.2%)
- other/uncertain	3,380	(2.3%)

4. Analysis

4.1. Extraction of code-mixing phenomena

In order to restrict our investigation to code mixing between the two major language choices, only transcription units showing an evident switch from (South Tyrolean) German to Italian or Trentino have been extracted from the corpus and subsequently analysed. This choice limits the scope of analysis in two ways: first, by excluding all switches occurring outside the boundaries of the utterance (or utterance sequence), i.e. between two clearly distinct utterances, conversational moves or turns (these phenomena are assumed as code switching); second, by considering only switches within the utterance containing surface elements from different codes, i.e. "matter" contact phenomena (cf. Matras and Sakel 2007)[6].

Another caveat concerns established borrowings: these phenomena are considered part of the lexicon of the receiving language and are not counted as switches [see (1)].

---

[5] ELAN is a software developed by *The Language Archive* (Max Planck Institute of Nijmegen); cf. Sloetjes and Wittenburg (2008). Excerpts from the corpus will be referenced with a code combining recording ID (e.g. "K016"), speaker's ID (e.g. "NG-17") and begin time (e.g. "5:37"; cf. example 1). Recordings with different types of text (interview, spontaneous speech, Map Task) were split into two or three transcription files (e.g. "K047A" and "K047B", cf. example 10), thus obtaining 61 ELAN files. Emphasis marks the switches; abbreviations are listed at the end of the article.
[6] A third obvious restriction is the exclusion of switches between Italian and Trentino, which, though present, will not be considered here. The two varieties are counted together in the following tables.

Similarly, proper nouns [see (2)][7], filled pauses and homophone interjections (such as *eh*, *mhm*) are not assigned to a specific language and do not constitute a proper switch.

Lastly, in a small group of cases the directionality of the switch is unclear [see (3)] due to the limited length of the utterance (2-3 items) and/or the kind of elements involved (e.g. interjections, discourse markers). These cases were excluded.

(1) gånz a grosä polenta
     really a big polenta:LW
     (K016.NG-17.5:37)

(2) wos see? see see lido?
     what lake … lido:PN
     (K003.P3-04.8:58)

(3) ah iå *beh*
     ah yeah *well*
     (K039.KT-Int.15:17)

4.2. Classification of code-mixed utterances

Data have been filtered from the corpus according to these principles, thus extracting 1,714 switches from German to Italian or Trentino and vice versa. Every switch has been categorized according to two surface parameters: directionality and extension. These two parameters are mutually dependent, often requiring that one be defined before the other can be observed.

Directionality is intended as the direction of the switch from the most activated code immediately before the switch point (code A) to the code activated by the switch itself (code B); in the analysis below, the directionality of the switch will be represented as "A>B".

The extension of the switch is the length of the switch in terms of the number of consecutive words in code B, allowing a first distinction between single-word switches and multi-word switches. Multi-word switches have been further analysed, observing their constituency as well as their dependency to elements outside the switch and highlighting diverging word order or structural patterns in the codes involved.

Table 2. Classification of code-mixing phenomena by directionality

Directionality of the switch	*N*	*%*
German > Italian/Trentino	1,457	85.0%
Italian/Trentino > German	257	15.0%
Total	*1,714*	

---

[7] Here *lido* refers to Bolzano's public outdoor swimming pool. It is used, untranslated, in German and English tourist information web pages (e.g. *Südtirol Info*), working as a proper noun for the specific location.

Table 3. Classification of code-mixing phenomena by extension

Extension of the switch	N	%
single-word switches	1327	77.4%
nested multi-word switches	197	11.5%
non-nested multi-word switches	190	11.1%
Total	1,714	

As Table 2 shows, there is a strong preference for switches from (South Tyrolean) German, which is the unmarked and dominant code for in-group interaction, to Italian or Trentino, which is often L2 for many bilingual speakers in South Tyrol. This strong asymmetry was expected and provides tangible evidence of the relationship between the codes in bilinguals' repertoire in South Tyrol.

Table 3 shows the classification of the switches according to their extension. Three different patterns have been identified:

A. *single-word switches* [see (4)]: all the switches involving a single isolated word or expression (e.g. *ma_dai* 'come on!' and other word combinations with a high co-occurrence frequency and/or discernible semantic autonomy), occurring in the middle or at the periphery of the unit (at the beginning or at the end, but still part of it), regardless of their lexical category or position in the argument structure;

B. *nested multi-word switches* [see (5)]: all multi-word switches formed by a single constituent or part of a constituent, with an evident structural dependency to the segment in the other code (thus, *selected* in the argument structure or *nested* in some other way, i.e. surrounded by structurally inter-dependent elements in code A);

C. *non-nested multi-word switches* [see (6)]: all multi-word switches formed of several constituents, thus representing a higher node in the syntactic tree and involving a potentially higher activation of code B from the switching point on, as well as single-constituent switches occurring at major clause boundaries and *non-nested*, i.e. not inserted within or surrounded by inter-dependent elements in code A or explicitly showing an improper word order in code A.

(4)   du gäa mål dain-ä *mult-a* zool-n
       you go.2Sg then your-F.Sg *fine-F.Sg* pay-INF
       'you go then and pay your fine'
       (K013.NG-09.10:21)

(5)   du muasch a *stori-a tragic-a* dårzeel-n
       you have_to.2Sg a *story-F.Sg tragic-F.Sg* tell-INF
       'you have to tell a tragic story'
       (K013.NG-Int.16:54)

(6)   haint hån i a *list-a* g-måch-t *con CENTO provinc-e italian-e*
       today have.1Sg I a *list-F.Sg* PP-make-PP *with one_hundred*
       province-F.Pl italian-F.Pl
       'today I made a list with a HUNDRED Italian provinces'
       (K013.NG-10.9:59)

As illustrated below, switches belonging to pattern A show a strong congruity with Muysken's class of insertion (see Section 4.3); on the contrary, the third class clearly

represents an alternational mixing; lastly, as it will be discussed in 4.4, multi-word switches of pattern B match insertional code mixing[8].

4.3. Single-word switches

The classification in Table 3 follows Muysken's (2000) model, precisely a subset of its diagnostic parameters[9], i.e. *single constituent*, *nested*, *selected* (denoting insertional mixing) and *several constituents*, *non-nested*, *major clause boundary* (which characterize alternation), except for one crucial issue: non-nested single-word switches. These are merged into the prototypical single-word insertion, which involves isolated content words in code B that are lexically activated and selected in the argument structure of an utterance in code A [see (4)].

Non-nested single-word switches, as shown in (7) and (8), typically involve adverbs, conjunctions, or discourse markers. In Muysken's model, these items can be interpreted as alternations (matching *adverb, conjunction* and/or *emblematic* or *tag* parameters). However, they do not share alternations' most characterizing and diagnostic features, i.e. *several constituents*, *long constituents*, *complex constituents* and *major clause boundary*.

Moreover, as Muysken himself points out, «[i]f we assume that nouns are borrowed through insertion and interjections through alternation, it is clear that insertional mixing is unidirectional and involves a matrix/non-matrix asymmetry, while alternational mixing is bidirectional» (Muysken 2000: 99). In our corpus, discourse markers switches share the same asymmetrical distribution as prototypical single-word insertions (cf. Dal Negro 2013).

(7)   unt når hån i gsåg *ma* sabine wiso net når håt si gsåg
       and then have.1Sg I say.PP *but* Sabine:PN why not then have.3Sg she say.PP
       'and then I said, "But, Sabine, why not?", then she said […]' (K017.NG-19.33:17)

(8)   häl glab=i *magari* war net bäas
       that believe.1Sg=I *maybe* be.COND.3Sg not bad
       'that - I think - maybe wouldn't be bad' (K001.P1-03.2:10)

A previous case study on the same data (Ciccolone and Dal Negro 2016) showed that the different functions performed by the same item (*ma*, 'but') activate different code-mixing patterns: while *ma* as discourse marker (alone or combined, as in *ma_dai*) shows an insertional pattern and is widely used by all speakers (even those using only German as base language in the whole corpus), original use of *ma* as adversative conjunction is a prerogative of speakers with a higher activation of Italian and/or Trentino and it systematically produces an alternation during the interaction.

Quantitative analysis on a subset of 1,000 code mixing phenomena extracted from the *Kontatto* corpus (involving a subset of speakers for whom we had more detailed sociolinguistic information) confirmed this relation between type (or function) of lexical items occurring as single-word switches and global activation of the code within the corpus. Moreover, distributional analysis of the subset showed the strong affinity to the insertional pattern of discourse-functional single-word switches [such as those in (7) and

---

[8] For further details and discussion on the classification of code-mixing phenomena and their relationship with insertional and alternational patterns, cf. Ciccolone and Dal Negro (2021).
[9] For a more detailed discussion of Muysken's parameters, see Ciccolone (2014).

(8)], as well as their higher frequency and diffusion among the observed speakers (cf. Dal Negro and Ciccolone 2018b).

These observations are in line with previous studies (cf. Fiorentini 2017; Dal Negro 2013, 2017; see also Matras 2009: 138ff. and 193ff.). All single-word switches have been thus identified with the insertional pattern.

4.4. Multi-word insertions

We will now focus on multi-word switches of single constituents or sub-constituents. These cases [see (9) and (11)] show remarkable congruence with the insertional pattern.

(9)   war schun *na figada*
        be.COND.3Sg really *a cool_thing*
        'it would be a really cool thing'
        (K022.KT-06.14:12)

(10)   main-ä *entrat-a nel mond-o metal dio_can* (K047A.AR-03.9:40)
        my-F.Sg *entry-F.Sg in_the world-M.Sg metal god_dog*
        'my introduction to the world of [heavy] metal - goddamnit!'

(11)   di martina sig i *'na volta ogni morte di papa*
        the Martina:PN see.1Sg I *one time every death of pope*
        'Martina, I see her once in a blue moon'
        (K026.KT-08.6:28)

While in (5) the Italian noun and its adjective are inserted in an NP with a German determiner, in (9) the determiner is also switched and the whole NP is inserted in the South Tyrolean utterance. (10) seems less obvious, but it still consists of a sub-constituent unit (a deverbal noun with a PP complement) inserted in an NP with a German specifier (the possessive pronoun).

(11) shows an Italian idiom inserted in a perfectly German construction (left dislocation) without any pause or disfluency: at its place we could have had either a frequency adverb (e.g. *selten*) or the corresponding German idiom (*alle Jubeljahre*).

It is interesting to notice that many other examples of multi-word insertions consist of an idiom or word combination with a certain degree of autonomy and loss of compositionality (cf. Bybee 2010: 44-50). Many of these word combinations fulfil specific discourse-structuring functions, such as:

– *general extenders* (example 12): they mark the end of a (potential) list and create an *ad hoc* category for the topic of discourse (cf. Overstreet 1999, Ajmer 2013);
– *closure markers* (example 13): they mark the end of a narrative sequence or the closing of a topic (cf. Drew and Holt 1998);
– *sum-up markers* (example 14): they mark the shift from informative to evaluative discourse (cf. Auer 1988: 199).

(12)   dår lino håt ålm di hörbüchår *e quelle robe là insomma* […]
        the Lino:PN have.3Sg always the audiobooks *and those things there so*
        'Lino has always the audiobooks and stuff like that'

(K047A.AR-04.25:13)

(13) /bam/ en colp scec via intel boschc *unt fertig*
      bam:MIM a blow hard away into_the wood *and ready*
      'Bam! a hard blow [and then] run into the woods and done!'
      (K022.KT-01.9:48)

(14) wail wänn dänk-sch # // *alla fine della fiera* isch es iå so
      because if think-2Sg # // *at_the end of_the fair* is it really so
      'because if you think # in the end it's so'
      (K033A.MT-05.6:10)

Table 4. Examples of discourse-structuring functions fulfilled by multi-word insertions

Function	*N*	*Examples*
General extenders	6	e quelle robe là, unt so (3), unt so waitår (2)
Closure markers	4	e basta, unt fertig
Sum-up markers	4	alla fine (3), alla fine della fiera

Another (also quantitatively) relevant subclass of multi-word insertions concerns compound lexical units like those in examples (15)-(16):

(15) draizän # *beh* mit *schpes-e di trasporto* sächzän
      thirteen # *well* with *fee-F.Pl of shipping* sixteen
      'thirteen # well, with shipping fees sixteen'
      (K013.NG-10.19:09)

(16) si måch-t glab=i irgendwo årbet-et si so in a *bancarella per frutt-i* wåasch?
      she make-3Sg believe.1Sg=I somewhere work-3Sg she so in a *stand*
      for fruit-M.Pl know.2Sg
      'I think she works somewhere at a fruit stand, you know?'
      (K028.KT-08.3:00)

Interestingly, the multi-word switches discussed here seem to follow the same patterns and functions of the two types of single-word insertions identified in Section 4.3, namely occasional, low-frequency content word insertions, with lexical-referential function on the one hand [see (4), (5), (15), (16)] and more frequent, peripheral insertions with a pragmatic (text-structuring or interactional) function on the other [see (7), (8), (13), (14)].

5. Discussion

In this study two surface parameters (directionality and extension) have been used to classify code-mixing phenomena and observe recurring patterns and functions in the combined use of codes within the same utterance or sequence of connected utterances. The distinction between single-word and multi-word switches proved to be crucial to determine the core features of the two different patterns of mixing observed: insertion and alternation.

Insertional mixing involves single isolated elements of code B embedded in an utterance otherwise well formed in code A and with a limited or occasional activation: (a) limited activation of a small set of (frequent) specific lexical items with a strong form-function univocity, mostly related to text and discourse structure (e.g. discourse and pragmatic markers); (b) occasional activation of random lexical items with a prominent referential function (content words such as nouns or adjectives).

This description of the insertional pattern applies to the whole set of single-word switches extracted from the corpus as well as to a group of multi-word switches with two diagnostic features: (a) they correspond to single constituents or sub-constituents; (b) they appear in the utterance without any evident violation of code A word order. These switches, here classified as multi-word insertions, fall into the same two macro-functional categories: (a) a limited set of pragmatic and textual functions realized by elements (combinations or idioms) of code B that gradually infiltrate utterances in code A; (b) occasional insertions of longer units (lexical compounds, noun-adjective pairs with high entrenchment, etc.) that allow bilingual users to manipulate a sort of "extended lexicon" by accessing the lexicon of code B.

The results of this analysis are in line with a previous prediction (cf. Ciccolone 2015: 76) on the distribution and frequency of insertional mixing phenomena, showing an exponentially higher number of single-word occurrences clustered according to their lexical category (with nouns, adjectives, and discourse-pragmatic markers at the top of the borrowability hierarchy, cf. Matras 2009; Poplack 2018: 48-50) and a decreasing number of occurrences in inverse relation with the extension of the switch. A thorough quantitative analysis shall be carried out to completely validate this prediction and to test it on different datasets concerning other language contact situations.

Abbreviations

–   1, 2, 3: first, second, third person
–   Sg, Pl: singular, plural
–   F, M: feminine, masculine gender
–   INF: infinitive
–   COND: conditional
–   PP: past participle
–   LW: attested loan word
–   PN: proper noun
–   MIM: mimetic (onomatopoeia)

References

Ajmer, Karin (2013), *Understanding Pragmatic Markers. A Variational Pragmatic Approach*, Edinburgh, Edinburgh University Press.

ASTAT (2017), *Statistisches Jahresbuch für Südtirol*, Bolzano, Provincia Autonoma di Bolzano/Alto Adige.

Auer, Peter (1988), 'A Conversation Analytic Approach to Code-Switching and Transfer', in Heller, Monica (ed.), *Codeswitching: anthropological and sociolinguistic perspectives*, Berlin-New York, de Gruyter, 187-213.

Auer, Peter (1999), 'From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech', *International Journal of Bilingualism* 3 (4), 309-332.

Berruto, Gaetano (2009), 'Confini tra sistemi, fenomenologia del contatto linguistico e modelli del *code switching*', in Iannàccaro, Gabriele; Matera, Vincenzo (eds.), *La lingua come cultura*, Novara, UTET, 3-34 and 212-216.

Berruto, Gaetano (2011), 'Interazione e compenetrabilità di grammatiche nel contatto linguistico. Il caso dell'enunciazione mistilingue', in Bombi, Raffaella; D'Agostino, Mari; Dal Negro, Silvia; Franceschini, Rita (eds.), *Lingue e culture in contatto. In ricordo di Roberto Gusmani*, Perugia, Guerra, 47-71.

Bybee, Joan (2010), *Language, Usage and Cognition*, Cambridge, Cambridge University Press.

Ciccolone, Simone (2014), 'Classificare il *code mixing*: una reinterpretazione dei parametri di *constituency* del modello di Muysken', *Linguistica e Filologia* 34, 95-134.

Ciccolone, Simone (2015), 'Interazione tra codici nel parlato bilingue: da fenomeni di contatto nel discorso all'emersione di schemi ricorrenti', in Consani, Carlo (ed.), *Contatto interlinguistico fra presente e passato*, Milano, LED, 53-81.

Ciccolone, Simone; Dal Negro, Silvia (2016), 'Marcare il contrasto nel parlato bilingue: *ma* e *obår* in un corpus sudtirolese', in Bombi, Raffaella; Orioles, Vincenzo (eds.), *Lingue in contatto/Contact linguistics*, Roma, Bulzoni, 97-113.

Ciccolone, Simone; Dal Negro, Silvia (2021), *Comunità bilingui e lingue in contatto. Uno studio sul parlato bilingue in Alto Adige*, Cesena, Caissa.

Dal Negro, Silvia (2013), 'Dealing with bilingual corpora: Parts of speech distribution and bilingual patterns', *Revue Française de Linguistique Appliquée* 18 (2), 15-28.

Dal Negro, Silvia (2017), 'Bilinguismo asimmetrico in Alto Adige: lo spazio sociolinguistico dell'italiano', in Bombi, Raffaella (ed.), *Nuovi spazi comunicativi per l'italiano nel mondo: l'esperienza di 'Valori identitari e imprenditorialità'*, Udine, Forum, 59-67.

Dal Negro, Silvia (2018), 'Finding patterns in bilingual speech', *Lingue e Linguaggio* 17 (1), 71-85.

Dal Negro, Silvia; Ciccolone, Simone (2018a), 'Il parlato bilingue: italiano e tedesco a contatto in un corpus sudtirolese', in Bermejo Calleja, Felisa; Katelhön, Peggy (eds.), *Lingua parlata. Un confronto fra l'italiano e alcune lingue europee*, Bern, Lang, 385-407.

Dal Negro, Silvia; Ciccolone, Simone (2018b), 'Incidenza ed estensione del *codemixing* come variabile sociolinguistica', *RID. Rivista italiana di dialettologia* 42, 119-144.

Deuchar, Margaret; Webb-Davies, Peredur; Donnelly, Kevin (2018), *Building and Using the Siarad Corpus. Bilingual conversations in Welsh and English*, Amsterdam-Philadelphia, Benjamins.

Drew, Paul; Holt, Elizabeth (1998), 'Figures of speech: Figurative expressions and the management of topic transition in conversation', *Language in Society* 27, 495-522.

Fiorentini, Ilaria (2017), *Segnali di contatto. Italiano e ladino nelle valli del Trentino-Alto Adige*, Milano, FrancoAngeli.

Matras, Yaron (2009), *Language Contact*, Cambridge, Cambridge University Press.

Matras, Yaron; Sakel, Jeanette (2007), 'Investigating the mechanisms of pattern replication in language convergence', *Studies in Language* 31 (4), 829-865.

Mioni, Alberto (1990), 'Bilinguismo intra- e intercomunitario in Alto Adige/Südtirol: considerazioni sociolinguistiche', in Lanthaler, Franz (ed.), *Mehr als eine Sprache*, Merano, Alpha & Beta, 13-35.

Muysken, Pieter (2000), *Bilingual Speech: A Typology of Code-mixing*, Cambridge, Cambridge University Press.

Overstreet, Maryann (1999), *Whales, Candlelight, and* Stuff *Like That: general extenders in English discourse*, New York, Oxford University Press.

Poplack, Shana (1980), 'Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching', *Linguistics* 18, 581-618.

Poplack, Shana (1987), 'Contrasting patterns of code-switching in two communities', in Wande, Erling; Anward, Jan; Nordberg, Bengt; Steensland, Lars; Thelander, Mats (eds.), *Proceedings from the Fourth Nordic Symposium on Bilingualism*, Uppsala, Borgströms, 51-77.

Poplack, Shana (2018), *Borrowing. Loanwords in the speech community and in the grammar*, Oxford, Oxford University Press.

Sloetjes, Han; Wittenburg, Peter (2008), 'Annotation by category – ELAN and ISO DCR', in *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008).

Südtirol Info = "Outdoor swimming pool", *Südtirol Info*, <https://www.suedtirol.info/en/experience/Outdoor-swimming-pool_activity_48504> (accessed 07/08/2021).

Simone Ciccolone
University of Cagliari (Italy)
simone.ciccolone@unica.it