Language Contact Phenomena in Catalonia:
The Influence of Catalan in Spoken Castilian

Mireia Galindo Solé
Universitat de Barcelona

1. Introduction[*]

Catalonia is a multilingual community, with two predominant languages: Catalan, the country's historical and native language, and Castilian (or Spanish) the official language of the state. The use of Castilian is widespread in Catalonia due both to sociopolitical developments and to intense immigration from Castilian-speaking regions. Both languages are recognized as official in Catalonia.

The fact that Catalan and Castilian have been in continuous contact over centuries has certainly modified both languages. However, the influence has not been symmetrical: the Castilian influence on Catalan has been more intense and pronounced. This imbalance is reflected in the analysis of linguistic influences in the two languages. While many studies have focused on language contact phenomena in Catalan, little research has been carried out on the Castilian used in Catalonia. In spite of these imbalances in influences and research, the varieties of Castilian spoken in Catalonia have certain interlinguistic features which differentiate it from the Castilian habitually spoken in monolingual areas.

In this paper we shall present the most salient conclusions of a study that analysed the extent to which language contact has modified the varieties of Castilian spoken in Catalonia. The research combines qualitative and quantitative methodologies, and is carried out from a structural point of view. The material gathered is a set of informal interactions produced by subjects of elementary school age.

2. Delimiting and classifying phenomena of language contact

The field of language contact has been analysed from a variety of perspectives and there is no generally accepted theoretical framework. As a result, many different labels have been used to refer to phenomena of language contact. Since classical times, when the terms *barbarism* or *foreign word* were used, various terms have been proposed. However, most of them have derogatory connotations or fail to fully express the idea. One of the terms that is relatively widely accepted is *transcodic marker*, coined by G. Lüdi (1987: 2) and defined as: «(...) marques, dans le discours, qui renvoient d'une manière ou d'une autre à la rencontre de deux ou plusieurs systèmes linguistiques (calques, emprunts, transferts lexicaux, alternances codiques, etc.)».

The term *transcodic marker* covers an extensive range of phenomena which have been organised in typologies (see, among others, Auer 1990 and Boix 1993). Although these typologies differ from each other, in general two broad groups can be distinguished. First, the phenomena that involve *code-switching*, that is «the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems» (Gumperz 1982: 59). Code-switching implies alternate activation of more than one language and is only possible in bilingual speakers. Second, phenomena traditionally named *interferences*, i.e. phonetic, lexical, morphosyntactic or semantic material transferred from one language to another; these phenomena become integrated inside the recipient language and are available to monolingual speakers, not only to bilinguals. This study focuses on the latter group of transcodic markers, the *interferences*; the practice of *code-switching* will be examined in a later study.

Just as there is no agreement on terminology, no single classification system has been accepted by all researchers. A wide variety of classifications have been adopted since the early days of research into the field, all of which differ in terms of the criteria taken as starting-points. One system organizes phenomena of language contact according to three linguistic levels: phonetic, lexical and grammatical (morphological and syntactic). These levels are not watertight compartments; the limits between them are seldom well defined, and features are rarely distributed in categories. Nonetheless, many studies are based on these linguistic levels (for instance, Weinreich 1953, Payrató 1985, Overbeke 1976 and Baetens 1986).

From another perspective, Haugen (1950) classified language contact phenomena according to their process of production and established a basic distinction between *substitution* and *importation*. The former, which gives rise to *calques*, involves (primarily semantic) transformations of items in a certain language due to the influence of another. The latter, which gives rise to *borrowings*, takes the form of transfers of complete units from the donor language into the recipient language.[1] Several criteria have been proposed for distinguishing between borrowing and code-switching. Variationist approaches (for example Poplack 1990) argue that borrowings are morphosyntactically integrated into the recipient language, while code-switching does not present this kind of integration. In contrast, other approaches (e.g. Myers-Scotton 1992) consider that it is the frequency of appearance that differentiates between the two phenomena: items that appear very frequently are regarded as borrowings, whereas those that appear only occasionally are classified as code-switching. Vila (1996: 389-406) applied these criteria in a study of Catalan and Castilian, concluding that frequency of appearance alone distinguishes between borrowings and code-switching in these languages.

There are other classification systems as well, which I will mention only briefly. One adopts the level of social integration as its starting-point. This criterion allows Weinreich (1953: 49) to distinguish between interferences in speech and interferences in language. Payrató (1985: 59) makes use of the same aspect to differentiate between interferences used only by bilinguals, those used by the whole of population, those explicitly accepted as normative and those which have entered the language many years before. Overbeke (1976: 114) also considers this criterion when distinguishing between interferences as a process and interferences as a result. Other classification systems take into account the relation between the interfered and interfering structure. Weinreich 1953 analyses this relation at the phonic level; Payrató 1985 extends it to other linguistic levels. Other classifications distinguish between positive and negative transfers, segmental and suprasegmental interferences, proactive and retroactive interferences, etc.

Some ways of classifying the results of language contact are summarized here.

Criteria	Classification systems
1. Linguistic level	→ phonic, lexical and morphosyntactic language contact phenomena
2. Process	→ substitution and importation
3. (Lack of) Integration into the recipient language	
3.1. Frequency of appearance 3.2. Morphosyntactic integration	→ borrowing and code-switching
4. Integration process stages	→ interference in speech and in language → interference as process and as result → etc.
5. Relation between interfered and interfering structures	→ under-differentiation, over-differentiation, reinterpretation, substitution, import and loss
6. Etc.	

Table 1. Some criteria and classification systems of language contact phenomena

---

[1] In Weinreich (1953) the process of importation is named *transfer*, and that of substitution is termed *interlinguistic identification*.

3. Language contact phenomena in Catalonia

In Catalonia, a wide range of language contact phenomena is detected in both Catalan and Castilian speech. It is often said that these phenomena appear frequently, but this is not in fact the case. Despite their relative scarcity in both languages, some differences between Catalan and Castilian speech have been observed: the percentage of transcodic markers in Catalan is significantly higher than those recorded in Castilian.

Vila's (1996) doctoral thesis analysed linguistic productions of children attending school in the Barcelona Metropolitan Area. The study showed that the amount of transcodic markers is comparatively low in both languages, though they were more frequent in Catalan: in Vila's corpus, transcodic markers constitute 3.57% of words included in Catalan speech, but only 0.57% in Castilian. Vila concluded that, although Catalan has been adopted as the main language of the educational system,[2] language contact phenomena in the Castilian used by the subjects of the study were rare. His results challenged the widely held assumption that the presence of Catalan in the school has contributed to introducing transcodic markers in the Castilian spoken by pupils.

4. The framework of the study
4.1. Goals and hypotheses

The major goal of this study was to analyse and quantify the language contact phenomena that appear in the Castilian spoken by children in Catalonia. The research focused on lexical and morphosyntactic transcodic markers. Due to logistical limitations, phonetic transcodic markers were not taken into account. My main hypothesis was that, even though pupils have a great deal of contact with Catalan in school, the influence of Catalan on the colloquial Castilian spoken by the children would be low.

4.2. Sample and methodological procedures

The research was based on a corpus of informal interactions, recorded during playtime in schools in several towns and cities in Catalonia. In order to guarantee spontaneity, the data were gathered semi-surreptitiously (some subjects were asked to carry a tape recorder hidden in their clothes). Forty-five samples of 15 minutes were obtained, amounting to 11 hours and 15 minutes of conversations. The subjects were children between the ages of 11 and 12, who were finishing primary education. All pupils knew both languages, although the degree of bilingualism varied according to family language: some subjects were native Castilian speakers, others native Catalan speakers, and others were from bilingual families.

Tape recordings were digitalized and transcribed in accordance with the transcription conventions of the Department of Catalan Philology (Universitat de Barcelona), proposed by Payrató (1995). Language contact phenomena were detected and coded at a later stage.

We considered transcodic markers the phenomena recorded in Castilian speech that showed a clear influence from a second language and absent in descriptions of present-day Castilian. The reference books consulted were: for Castilian, Moliner (1966), Seco, Andrés and Ramos (1999), Seco (1986) and Alarcos (1994); for Catalan, Enciclopèdia Catalana (1987), Badia (1994), Solà (1994), and, to compare them, Enciclopèdia Catalana (1998). However, among the phenomena we considered as transcodic markers were items and structures included in descriptions as peculiar to Catalan-speaking areas (in section 5, these forms are marked with an asterisk), and also phenomena showing a clear Catalan influence and described in reference books as *archaisms* (habitual features in old Castilian, but no longer used today).

---

[2] Between 1978 and 1983, Catalan was a compulsory subject in school, taught for three hours a week. Between 1983 to 1994 it became compulsory to teach some subjects in Catalan. Today Catalan is the main vehicle of instruction (Vila 2000).

The phenomena detected were classified as lexical or morphosyntactic transcodic markers. Lexical items were categorized according to their process of formation, enabling us to distinguish between loanwords and calques. The criteria "frequency of appearance" and "inclusion in earlier studies" were used to distinguish between widespread loanwords and calques on the one hand and code-switches and non-widespread calques on the other: items appearing in the speech of more than two children or mentioned in previous studies were regarded as borrowings or widespread calques; those produced by one or two subjects and not previously reported elsewhere were considered one word code-switches or non-widespread calques. Phenomena of *convergence* were analysed separately. This label covers features of Castilian which are attributed a wider frequency of use in Catalan-speaking areas than in others, due to coincidence with their Catalan equivalents.

5. Data analysis: language contact phenomena in Castilian
5.1. Some examples of convergence

Before presenting the transcodic markers detected in the corpus, we will analyse a number of features which, though not unknown in the speech of Castilian speakers in monolingual areas, are believed to be more frequent in Catalonia due to the fact that they coincide with Catalan forms. We know of no quantitative studies that corroborate the differences in frequency of use, and so the results obtained must be considered provisional.

One of the most commonly cited examples of convergence between Catalan and Castilian is the use of definite articles before proper names (example 1*a*, *b* and *c*). In Catalan, the phenomenon is widespread in colloquial and standard language, although it tends to be avoided in formal speech. In Castilian it is frequent too; however, the use of the personal article is considered a vulgar feature, typical of the lowest sociolects. The use of the personal article in the corpus is practically constant, as shown in table 2. It should therefore be considered a distinctive feature of varieties of Castilian spoken in Catalonia.

(1) *a.*   SAB: profe_ *el* Albert ha cola(d)o la pelota\
              (*teacher, Albert has kicked the ball out of the playground*)

    b.   VAN: después de *la* Desi yo\
              (*after Desi, me*)

    c.   CAR: a mí me ha toca(d)o con *el* Brian y *el* Brad Pitt\
              (*I've got Brian and Brad Pitt*)

linguistic features	absolute	relative
article + proper name	552	.942
Ø + proper name	34	.058
Total	586	1

Table 2. Frequencies of definite articles before proper names.
Absolute and relative figures

Another example of convergence affects deixis. In Castilian, deictics are structured in a ternary system: the first person covers the immediate context to speaker, the second person alludes to the interlocutor, and the third person refers to other people. Like English, Catalan uses the first person, which represents the space near the speaker, and the second person, representing distance from the speaker (see table 3).

	Catalan (English)		Castilian		
	1st and 2nd p.	3rd p.	1st p.	2nd p.	3rd p.
locatives	*aquí* (*here*) –	*allí/allà* (*there*)	*aquí* –	*ahí* –	*allí/allá*
demonstratives	*aquest* (*this*) –	*aquell* (*that*)	*este* –	*ese* –	*aquel*
neuter pronouns	*això* (*this*) –	*allò* (*that*)	*esto* –	*eso* –	*aquello*

Table 3. Deictic system. Differences between Catalan, Castilian and English

The three-level system is often reduced to two levels in colloquial Castilian (Alcina and Blecua 1975: 622). This tendency may be more noticeable in Catalonia owing to the coincidence with the Catalan system. The analysis of deictics always depends on the interpretation of the words of an informant, and so the researcher may misunderstand the speaker's communicative aims. For this reason, they have not been quantified here. Nevertheless, some of the cases of this transformation compiled during the study are shown in the following examples. In (2*a*) JOS uses *esto* to reproduce an affirmation made by his interlocutor, and not by himself; in (2*b*) MAR uses *este* (and no *ese* or *aquel*) to refer to a subject that is not present at that particular moment.

(2) *a.* CAR: nunca la lleva_ eh/
       JOS: bueno: pero *esto* es normal\
             (*he never wears it / well, this is normal*)
     b. ANN: qué te pasa\
       MAR: el niño *este*_ que es tonto\
       ANN: quién_ el Roger_
             (*what's the matter? / this boy is silly / who? Roger?*)

Phenomena of convergence also affect verbal morphology. The use of *ves*, instead of *ve* (second person singular imperative of *ir* 'go'), has been reported as a vulgar feature throughout Spain (Seco 1986: 236). However, it seems highly likely that the existence of the Catalan equivalent *vés* has increased the use of *ves* in Catalonia. Due to the extension phenomenon (shown in table 4), *ves* can be considered the habitual form in Catalonia, at least in the age group studied (this feature is exemplified in 3).

(3)   NKI: sí: sí_ tú *ves* detrás del Raúl_
                (*yes, yes, you go behind Raúl*)

linguistic features	absolute	relative
ves	7	.875
ve	1	.125
total	8	1

Table 4. Frequencies of *ves and ve* (second person singular
imperative of '*ir*'). Absolute and relative figures

Using the possessive after the preposition instead of a prepositional phrase (the normative form) is habitual in certain registers of language (Moliner 1966, Alarcos 1994). This construction may be more frequent in Catalonia because it coincides with Catalan equivalents. The corpus contains more cases of prepositions followed by a possessive (example 4) than of prepositions followed by a prepositional phrase. However, given the similarity in occurrences of the variants and the low frequency of appearance of this variable we cannot regard this construction as the general form in the Castilian spoken in Catalonia (see table 5).

(4)   LYD: es que-- verdad que la Merche iba *detrás mío*_
                (*Merche was behind me, wasn't she?*)

linguistic features	absolute	relative
de + pronoun	4	.4
possessive	6	.6
total	10	1

Table 5. Frequencies of preposition + prepositional phrase
and preposition + possessive. Absolute and relative figures

Convergence also involves other constructions. *Es veu que*, meaning 'it seems (to be)', is habitual in colloquial Catalan. In Castilian, two expressions adopt this meaning: *se ve que* and *por lo visto*. However, *por lo visto* would be expected to be more frequent in this context, whereas *se ve que* is generally used to mean *one observes that* (Hernández 1998: 188). No examples of *por lo visto* were recorded in our corpus, but two occurrences of *se ve que* 'seem (to be)' were detected, its appearance perhaps favored by the Catalan form (5).

(5)   AKU: se ve que se ha comido un tocho_ entonces_
               (*it looks as if he ate a brick, then*)

The literature mentions other examples of potential convergence: *se ha(n) de* instead of *hay que* in impersonal periphrasis of obligation, the use of *haber de* in place of *tener que* in personal periphrasis of obligation, or a tendency to introduce the second part of a comparison with *que no*. The data analysis does not confirm the frequency of appearance of structures that coincide with the Catalan ones. Table 6 shows that the variants that are more frequent in non Catalan-speaking areas are also the habitual forms in the varieties of Castilian spoken in Catalonia.

linguistic features	absolute	relative
impersonal periphrasis of obligation	9	1
hay que	9	1
se ha de	0	0
personal periphrasis of obligation	104	1
tener que	103	.990
haber de	1	.010
comparisons with negation	8	1
que	7	.875
que no	1	.125

Table 6. Frequencies of some linguistic features.
Absolute and relative figures

We were unable to study the frequency of other phenomena generally attributed to convergence towards Catalan, since these variables presented insignificant or non-existent occurrence ratings. Examples are the use of *deber* instead of *deber de* to express probability, a tendency which is believed to be higher in the Castilian spoken in Catalonia because the Catalan equivalent *deure* (*deber*) is the only form used to express probability in that language. Nor could we analyse the extent to which there was agreement between certain impersonal verbs (such as *hacer* or *haber*) and following plural nouns.

5.2. Lexical contact phenomena

Transcodic markers due to enlargement, restriction or shift of the meaning are not very frequent (see table 7). All the calques recorded (excepting *piel de gallina*) have been reported in previous research; for this reason they are classified as widespread.

Here we will only present two of the examples recorded. The first is the use of *venir* instead of *ir*. This feature has been regularly cited in the literature as a characteristic of the Castilian spoken in Catalonia. In Castilian, *ir* means motion only away from the speaker, while *venir* signifies motion towards the speaker. In Catalan, in contrast, *venir* covers both meanings. Example 6 shows AKS uses *venir* to tell ANA that she is going to go towards her.

(6)   AKS: {(AC) qué estás haciendo\}
       ANA: aquí:_ jugando:_ en el parque\
       AKS: ahora *vengo*\
               (*what are you doing? / here, I am playing in the park / I am coming*)

lexical transcodic markers (calques)	absolute frequencies	relative frequencies	Catalan equivalent	Castilian equivalent	English equivalent
widespread calques					
a más a más	1	.1	*a més a més*	*además*	*moreover*
hacer campana	1	.1	*fer campana*	*hacer novillos*	*to play truant*
hacer miedo	1	.1	*fer por*	*dar miedo*	*to scare*
ser	1	.1	*ser*	*estar*	*to be*
suerte que	3	.3	*sort que*	*menos mal que*	*it is just as well that*
venir	2	.2	*venir*	*ir*	*go*
non-widespread calques					
piel de gallina	1	.1	*pell de gallina*	*carne de gallina*	*gooseflesh*
	10	1			

Table 7. Widespread and non-widespread calques. Absolute and relative frequencies.
Catalan, Castilian of non Catalan-speaking areas and English counterparts.

The varieties of Castilian spoken in Catalonia are also characterized by the use of the verb *hacer* in a multiplicity of contexts in which standard Castilian adopts other forms. The only example recorded in the corpus was *hacer miedo* in place of *dar* or *meter miedo (in English, to frighten someone)*; nevertheless in the Castilian spoken in Catalonia it is common to hear *hacer tarde* (instead of *llegar tarde, in English, to be late*), *hacer un café* instead of *tomar un café, (in English, to have a cup of coffee)*, etc. (Casanovas 1996).

The transcodic markers that we recorded entered the recipient language more frequently via importation than via substitution. In all, 27 types of loanwords and code-switches were recorded, compared with 7 types of calques. This difference was also observed with regard to tokens: 91 vs. 10 (compare tables 7 and 8).

As far as types of loanwords and code-switches are concerned, the corpus was clearly biased towards a number of specific domains, for instance, the semantic field of games and leisure (*xarranca*, *cau*, *esplai*, *amagar*), school equipment (*estijeras*), names to refer to teachers (*professor*) and terminology from specific subject areas (*vat*).

The following examples present the use of borrowings. Example 7 illustrates the use of *nen* (child, instead of *niño* or *tío*), a widely used discourse marker in this age group. Tokens of *cau* and *esplai* are shown in 8 (these words refer to centers where children go at the weekend or on holiday and participate in activities such as excursions, handicraft, etc.). Example 9 presents the use of *vats*. In fact, in the next speech turn a native bilingual corrected NKC, and so *vats* is taken to be a non-widespread transcodic marker.

(7) RUB: {(F) ah:_ la mano tío_} la mano_ *nen*_
      (*ah, my hand, man, my hand, man*)

(8) *a.* PAT: a la Laura se lo contaron en el *cau* también_
      (*someone in the play center told Laura*)
  b. SAR: {(??) porque como no tengo esplai_}
      (as I'm not going to the play center)

(9) NKC: y el A- el Aníbal tiene cuarenta *vats* y no puede sacar las chincheta(s)_
    RIC: {(@) cuarenta vatios dice\}
      (*Aníbal has forty watts and he cannot take out the thumbtacks*
      / *forty watts he says!*)

lexical transcodic markers (loanwords and switches)	absolute frequencies	relative frequencies	Catalan equivalent	Castilian equivalent	English equivalent
loanwords					
*amagar**	7	.0769	*amagar*	*esconder*	*to hide*
ara	15	.1648	*ara*	*ahora*	*now*
caldre	1	.0110	*caldre*	*hacer falta*	*to be lacking*
cau	2	.0220	*cau*	---	---
coca	1	.0110	*coca*	*torta, bizcocho*	*cake*
esplai	1	.0110	*esplai*	---	---
nen	7	.0769	*nen*	*niño, chico*	*boy*
va	29	.3187	*va*	*venga*	*come on*
pencar	2	.0220	*pencar*	*currar*	*to work*
tocho	2	.0220	*totxo*	*ladrillo*	*brick*
one word switches					
cagarro	1	.0110	*cagarro*	*mojón*	*idiot*
*chafardero**	1	.0110	*xafarder*	*chismoso*	*gossiping*
conversa	1	.0110	*conversa*	*conversación*	*talk*
escañarse	1	.0110	*escanyar-se*	*atragantarse*	*to choke*
esquerda	1	.0110	*esquerda*	*grieta*	*crack*
estijeras	1	.0110	*estisores*	*tijeras*	*pair of scissors*
farigola	3	.0330	*farigola*	*tomillo*	*thyme*
infermera	1	.0110	*infermera*	*enfermera*	*nurse*
noia	1	.0110	*noia*	*chica*	*girl*
porqué(t)	6	.0659	*porquet*	*cerdito*	*little pig*
professor	1	.0110	*professor*	*profesor*	*teacher*
reixa	1	.0110	*reixa*	*reja*	*gate*
tercer	1	.0110	*tercer*	*tercero*	*third*
vat	1	.0110	*watt*	*vatio*	*watt*
vinga	1	.0110	*vinga*	*venga*	*come on*
xarranca	1	.0110	*xarranca*	*calderón*	---
xurisu	1	.0110	*xoriço*	*chorizo*	*hard pork sausage*
	91	1			

Table 8. Loanwords and one-word switches. Absolute and relative frequencies.
Catalan, Castilian of non Catalan-speaking areas and English counterparts.

5.3. Morphosyntactic contact phenomena

Castilian displays fewer innovative features at the morphosyntactic than at the lexical level. The figures in tables 7, 8 and 9 show these differences.

The above table shows that two examples of pronominalization were detected: both *adelgazar* 'to slim' and *pensar* (when it means 'to believe') are pronominalized, following the characteristics of their Catalan counterparts (10 *a* and *b*).

(10) *a.*   EST: {(F) por tu culpa *se piensan* que tú y yo somos novios\}
              (*because of you, they think that you and I are a couple*)
    b.   POL: y *te adelgazas*\
              (*and you lose weight*)

morphosyntactic transcodic markers	absolute frequencies	relative frequencies	Catalan equivalent	Castilian equivalent	English equivalent
pronominalization					
adelgazarse	1	.0263	*aprimar-se*	*adelgazar*	*to slim*
pensarse	5	.1316	*pensa-se*	*pensar*	*to believe*
confusion between prepositions					
a instead of *en*	7	.1842	*a*	*en*	
a instead of *por*	1	.0263	*a*	*por*	
con instead of *en*	1	.0263	*amb*	*en*	
de instead of *a*	1	.0263	*de*	*a*	
por instead of *para*	2	.0526	*per*	*para*	
introduction of expletive *que* in interrogatives	4	.1053	que	Ø	Ø
deletion of *a* to introduce direct object	6	.1579	Ø	*a*	Ø
introduction of expletive *que* in exclamations	8	.2105	que ... que	qué ...	how ... Ø
tampoco no	1	.0263	tampoc no	tampoco	not ... either
	38	1			

Table 9. Absolute and relative frequencies of morphosyntactic transcodic markers.
Catalan, Castilian of non Catalan-speaking areas and English counterparts.

Contact phenomena also included confusion in the use of prepositions, due to non-coincidence between Castilian and Catalan systems. The preposition *a* is an example. In Catalan this preposition introduces a place complement (place where); in Castilian this function is assigned to the preposition *en* (11). There was also confusion between *con* and *en*: in Castilian, means of transport are introduced by *en*, whereas in Catalan both *en* and *amb* (*con*) are correct (12). The last example illustrates the use of *por* instead of *para*. In colloquial Catalan, *per* (*por*) covers all the functions that in Castilian are shared between *por* and *para*. Because of the influence of Catalan, the sample's Castilian is biased in favor of a single preposition for all meanings, generally *por*. In example 13 *por* is used instead of *para* to express comparison or disproportion.

(11)   XIO: {(F) que no vamos a meternos *a*l agua\}
                         (*we won't go in the water*)
(12)   NKD: por la: por las tardes_ iremos a la xxx_ *con* las motos_
                         (*in the afternoon we will go to xxx by motorbike*)
(13)   SUS: *por* un día que salimos_ ya os tenéis que cabrear_ o qué\ colegas\
                         (*we only go out one day and then you get angry*)

Another feature that characterizes the Castilian spoken in Catalonia is the use of interrogative expletive particle *que* (14). Interrogatives introduced by *que* which reproduce Catalan structures should not be confused with those which are possible in Castilian. The former present a final falling tone (*¿Que vais a usar mono?*), in the latter the pitch accent rises (*¿Cómo? ¿Qué dices? ¿Que vais a usar mono?*).

(14)   RUB: *que* vais a usar mono_
           EST: [no:_]
                        (*are you going to you use overalls? / no*)

Other language contact phenomena at the morphosyntactic level are: placing the expletive particle *que* in front of the second part of exclamations (15); omitting the preposition *a* before person direct object (16) — this omission is required in Catalan, although it is not always executed in colloquial speech; and placing the adverb *no* after the adverb *tampoco* (17). These phenomena have been

included in this section although their real nature remains doubtful: it has not been established whether they are exclusive to the Castilian spoken in Catalonia.

(15)   REB: qué pesa(d)os *que* sois_
         (*you are very annoying*)
(16)   ELE: {(??) voy a ver Ø mi niño_}
         (*I am going to see my baby*)
(17)   NE1:   pero *tampoco no* encontraron_
         (*but they didn't find that either*)

5.4. Convergence and transcodic markers in the corpus

Transcodic markers represent a very low proportion of the amount of words in the corpus, both in terms of lexical contact phenomena (0.25%, table 10) and morphosyntactic phenomena (0.09%, table 11). Taking all transcodic markers together, the percentages remain low (out of a total of 39941 items, only 139 tokens were recorded, i.e. 0.35%, table 12).

	tokens	%
number of lexical TM	101	0.253
number of words (without lexical TM)	39840	99.747
Total number of words	39941	100

Table 10. Number of lexical transcodic markers over total number of words. Absolute and relative frequencies

	tokens	%
number of morphosyntactic TM	38	0.095
number of words (without morph. TM)	39903	99.905
Total number of words	39941	100

Table 11. Number of morphosyntactic transcodic markers over total number of words. Absolute and relative frequencies

	tokens	%
number of TM	139	0.348
number of words (without TM)	39802	99.652
Total number of words	39941	100

Table 12. Number of transcodic markers over total number of words. Absolute and relative frequencies

The proportion of convergence phenomena was slightly higher: out of a total of 39941 words, the study detected 571 items (i.e., 1.43%) which may be regarded as more frequent in varieties of Castilian spoken in Catalonia (table 13).

	tokens	%
number of convergence phenomena	571	1.430
number of words (without conver.)	39370	98.570
Total number of words	39941	100

Table 13. Number of convergence phenomena over total number of words. Absolute and relative frequencies

Adding up the items analysed does not significantly increase the percentages: only 1.78% of words included in the corpus could be considered the consequence of the influence of Catalan, due to

either importation or substitution, or because of convergence to Catalan (table 14). The results shown in tables 13 and 14 provide quantitative evidence that supports the research hypothesis.

	tokens	%
convergence + transcodic markers	710	1.778
number of words (without conv. + TM)	39239	98.222
Total number of words	39941	100

Table 14. Number of convergence phenomena and transcodic markers compared with total number of words. Absolute and relative frequencies

6. Conclusions

The main goal of this study was to determine the extent to which the language contact situation in Catalonia has modified Castilian spoken in this area. Data analysis reveals that phenomena of language contact represent a low percentage in the speech of the subjects. Lexical and morphosyntactic transcodic markers constitute 0.35% of the total of the words examined in the corpus. If convergence phenomena are added, the percentage increases to 1.78%, still a low figure. All of the results presented here lead to the same conclusion: the influence of Catalan on the Castilian spoken in Catalonia is minimal. Considering the composition of the sample (pupils from a range of elementary schools around Catalonia), it can be said that the increase in the use of Catalan in education has not modified the Castilian spoken by children.

Lexical transcodic markers were more abundant and varied than morphosyntactic ones. As far as loanwords are concerned, *va* (*vamos*) was the most frequently recorded form (29 tokens), followed by *ara* (*ahora*, 15), *nen* (*chico*, 7) and *amagar* (*esconder*, 7). With regard to calques, the expression *suerte que* and the use of *venir* instead of *ir* were the features that presented more than one token (2 and 3 respectively). At the morphosyntactic level, confusion between prepositions was the most common phenomenon (12 tokens), followed by the introduction of the expletive *que* in exclamations (8), the pronominalization of some verbs (6), the deletion of the preposition *a* to introduce personal direct objects (6) and the use of the interrogative expletive particle *que* (4). As for the mechanisms that produce transcodic markers, importation was more frequent than substitution at the lexical level, while interlinguistic identification was the only process at the morphosyntactic level.

Two aspects must be borne in mind while interpreting these data: the nature of the data analysed, and the age of the speakers. The high number of tokens of *va* (*vamos*), for instance, can only be understood if we remember that is a recurrent expression in colloquial language, and rarely appears in formal register. Likewise, the frequency of appearance of *nen* is obviously related to the age of the subjects: this word would hardly ever be recorded in data produced by other generational groups. Moreover the quantitative results must be considered in the light of the circumstances of recording. The children were free to speak the language they chose, since researchers gave no instructions and the data were recorded without the presence of either teachers or researchers. We only examined Castilian utterances, which supposedly was produced by children who felt comfortable speaking this language. If all subjects had been asked to speak Castilian, the amount of language contact phenomena would have risen.

We should also bear in mind the delimitation and classification of language contact phenomena adopted. This was a synchronic study and so transcodic markers introduced many years previously were not analysed. We took recent descriptions of Castilian as our basis for detecting transcodic markers, and regarded frequent items and structures as proper to Castilian in spite of the fact that they are not included in prescriptive works. Another aspect that had a bearing on the final results is the classification system. Following the information available in reference works, we regarded *ves,* 'go', as convergence, although Vila (1996) regarded it as a morphological transcodic marker. For the same reasons we consider the omission of *a* before a direct object as a transcodic marker, though Hernández (1998) classified this feature as a convergence phenomenon.

References

ALARCOS LLORACH, Emilio (1994): *Gramática de la lengua española*. Madrid: Espasa-Calpe.

ALCINA, Juan; BLECUA, José Manuel (1975): *Gramática española*. Barcelona: Ariel.

AUER, P. (1990): «A discussion paper on code alternation». In NETWORK ON CODE-SWITCHING AND LANGUAGE CONTACT, 69-88.

NETWORK ON CODE-SWITCHING AND LANGUAGE CONTACT (1990): *Papers for the Workshop on Concepts, Methodology and Data. Basel, 12-13 January 1990.* Strasbourg: European Science Foundation.

BADIA I MARGARIT, Antoni (1994): *Gramàtica de la llengua catalana*. Barcelona: Enciclopèdia Catalana.

BAETENS BEARDSMORE, Hugo (1986): *Bilingualism: Basic Principles*. Clevedon: Multilingual Matters. 2nd ed.

BOIX, Emili (1993b): *Triar no és trair. Identitat i llengua en els joves de Barcelona.* Barcelona: Edicions 62.

BOIX, Emili; VILA, F. Xavier. (1998): *Sociolingüística de la llengua catalana*. Barcelona: Ariel.

CASANOVAS, Montserrat (1996): «Algunos rasgos propios del español en las comunidades de habla catalana: fonética, morfosintaxis y léxico». *Analecta Malacitana. Revista de la Sección de Filología de la Facultad de Filosofía y Letras*, XIX, 1, 149-160.

ENCICLOPÈDIA CATALANA (1987): *Diccionari català-castellà / castellà català*. Barcelona: Enciclopèdia Catalana.

ENCICLOPÈDIA CATALANA (1998): *Gran diccionari de la llengua catalana*. Barcelona: Enciclopèdia Catalana.

HAUGEN, Einar (1950): «The analysis of linguistic borrowing». *Language*, 26, p.210-231.

HERNÁNDEZ, Carmen (1998): *Algunas cuestiones más sobre el contacto de lenguas: Estudio de la interferencia lingüística del catalán en el español de Cataluña*. Unpublished Ph.D. Thesis. Universitat de Barcelona.

LÜDI, Georges (1987): «Les marques transcodiques: regards noveaux sur le bilinguisme». In Devenir bilingüe - parler bilingüe. Actes du 2e colloque sur le bilinguisme, Université de Neuchâtel, 20-22 septembre 1984. Tübingen: Niemeyer, 1-19.

GUMPERZ, John Joseph (1982): *Discourse Strategies*. Cambridge: Cambridge University Press.

MOLINER, María (1966): *Diccionario de uso del español*. Madrid: Gredos.

MYERS SCOTTON, Carol (1992): «Comparing codeswitching and borrowing». *Journal of Multilingual and Multicultural Development*, 13, 19-39

OVERBEKE, Maurits van (1976): *Mécanismes de l'interférence linguistique*. Madrid: Fragua.

PAYRATÓ, Lluís (1985): *La interferència lingüística. Comentaris i exemples català-castellà*. Barcelona: Curial; Publicacions de l'Abadia de Montserrat.

PAYRATÓ, Lluís (1995): «Transcripció del discurs oral». In PAYRATÓ *et al*. (ed.): *Corpus, corpora*. Barcelona: PPU; Departament de Filologia Catalana, 181-216.

POPLACK, Shana (1990): «Variation theory and language contact: concepts, methods and data». In NETWORK ON CODE-SWITCHING AND LANGUAGE CONTACT, 34-62.

SECO, Manuel (1986): *Diccionario de dudas y dificultades de la lengua española*. Madrid: Espasa-Calpe.

SECO, Manuel; ANDRÉS, Olimpia, RAMOS, Gabino (1999): *Diccionario del español actual*. Madrid: Aguilar.

SOLÀ, Joan (1994): *Sintaxi normativa: estat de la qüestió*. Barcelona: Empúries.

VILA I MORENO, F. Xavier (1996): *When Classes Are Over. Language Choice and Language Contact in Bilingual Education in Catalonia*. Unpublished Ph.D. Thesis. Vrije Universiteit Brussel.

VILA I MORENO, F. Xavier (2000): «Les polítiques lingüístiques als sistemes educatius dels territoris de llengua catalana». *Revista de Llengua i Dret* 34, 169-208.

WEINREICH, Uriel (1953): *Languages in Contact. Findings and Problems*. The Hauge-Paris: Mouton.

Selected Proceedings of the First Workshop on Spanish Sociolinguistics

edited by Lotfi Sayahi

Cascadilla Proceedings Project   Somerville, MA   2003

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document #
which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Galindo Solé, Mireia. 2003. Language Contact Phenomena in Catalonia: The Influence of Catalan in Spoken
Castilian. In *Selected Proceedings of the First Workshop on Spanish Sociolinguistics*, ed. Lotfi Sayahi, 18-29.
Somerville, MA: Cascadilla Proceedings Project.

or:

Galindo Solé, Mireia. 2003. Language Contact Phenomena in Catalonia: The Influence of Catalan in Spoken
Castilian. In *Selected Proceedings of the First Workshop on Spanish Sociolinguistics*, ed. Lotfi Sayahi, 18-29.
Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1004.