

## ON THE RELATIONSHIP BETWEEN COMPREHENSION AND PRODUCTION DATA IN CODESWITCHING

PAOLA E. DUSSIAS

*The Pennsylvania State University*

### 0. *Introduction*

Numerous studies examining the regularity of occurrence of types of codeswitches in bilingual production data reveal that there is a systematic favoritism for switches that involve certain grammatical categories over others during codeswitched speech (see, for example, DiSciullo, Muysken & Singh 1986, Halmari 1997, Klavans 1985, Lance 1975, Lipski 1985, Myers-Scotton 1993, Park, Troike & Park 1993, Poplack 1980, Sankoff & Poplack 1981, Treffers-Daller 1995). In this respect, the available literature indicates that whereas functional elements tend to appear in one language, their complements appear in the other language (Muysken 1997). This phenomenon has come to be known in the literature as ‘the functional element effect.’<sup>1</sup> Several studies looking at Spanish-English codeswitching data, for example, have shown that when switches involve nominal phrases with a Determiner (Det) and a Noun (N), instances in which both Det and N appear in the same language occur less frequently than cases in which Det appears in one language and N in the other. This asymmetry has been attested in Sankoff & Poplack (1981), who note that the relative propensity of switches occurring between a determiner and a noun was 13% in their data, compared to the 2.3% figure representing codeswitched sentences containing the determiner and the noun in the same language. Another example is provided in a recent study by Milian (1996, cited in Myers-Scotton & Jake 1997), who reports that 81% of the switches involving determiners and their

---

<sup>1</sup> Following Cowper (1992), I define functional elements as categories that lack substantive meaning, do not assign theta roles, are closed classes (no new words can be created), and do not permit recursion on X-bar. Lexical elements, on the other hand, are defined as categories that have substantive meaning, assign theta roles to their arguments, are open classes, and permit indefinite recursion on X-bar. Accordingly, COMP, INFL, and DET are instances of functional elements, and N, V, and A are instances of lexical elements.

corresponding nouns consist of English nouns and Spanish determiners, but only 6% of the switches involve English determiners accompanied by English nouns.

The limited participation of Complementizer (Comp) in codeswitching provides another example of this functional element effect. To take an example, Sankoff & Poplack (1981) report the propensity for a subordinate conjunction to be the locus of a codeswitch at less than 0.2% in their data, which contrasts with the 3.9% representing the propensity of switches that occur at the category following the subordinate conjunction<sup>2</sup>. In addition, the observation that subordinate conjunctions preferentially appear in the language of the head element on which they depend (i.e., what DiSciullo, Muysken & Singh 1986, Halmari 1997, and Treffers-Daller 1995 refer to as ‘the governor’ or ‘case assigner’) suggests that complementizers are subject to minimal participation in codeswitching. Examples (1) through (4) illustrate the types of codeswitches involving Det-NP and Comp-IP clauses most frequently encountered in naturalistic data:

- (1) *Mis padres van a venir para los holidays.*  
 My parents are going to come for the holidays.  
 “My parents are going to come for the holidays.”
- (2) *Es una little box asina y ya viene...*  
 Is a small box like this and already comes...  
 “It is a small box like this and it comes already ...”
- (3) *No, la potato de anoche, you acabaste con ella.*  
 No, the potato from last night, you finished with it.  
 “No, the potato from last night, you finished it.”
- (4) *I’m not saying that son chuecos, yo no digo eso.*  
 I’m not saying that are old, I NEG say that.  
 “I’m not saying that they are old, I am not saying that.”

---

<sup>2</sup> See Lipski (1985) for similar results.

### 1. *The functional element effect in comprehension data*

In a series of studies, Dussias (1997, 1999, forthcoming) investigated whether the functional element effect, often observed during sentence production, occurred in comprehension as well. Dussias (1999), for example, conducted two experiments in which subjects' eye-movements were recorded while reading sentences with codeswitches between functional heads and their complements. Twenty-four Spanish-English bilinguals participated in the study. All subjects had learned the two languages before the age of six and reported using both languages in their daily lives and in a variety of contexts.

Experiment 1 investigated codeswitches involving Comp and its IP complement by looking at reading performance in two conditions. In Condition 1, the codeswitch occurred at Comp and in Condition 2, it occurred at IP. A sample of each condition is given in (5) and (6) below:

- (5) *La maestra no sabía that the boy had left.*  
 The teacher neg know that the boy had left.  
 "The teacher did not know that the boy had left."
- (6) *La maestra no sabía que the boy had left.*  
 The teacher neg know that the boy had left.  
 "The teacher did not know that the boy had left."

Experiment 2 investigated switches that involved Det and its complement NP. Again, two conditions were compared. In Condition 1, the codeswitch occurred at Det, whereas in Condition 2 the switch occurred at NP. This is illustrated in (7) and (8):

- (7) *La maestra compró the books for the children.*  
 The teacher bought the books for the children.  
 "The teacher bought the books for the children."
- (8) *La maestra compró los books for the children.*  
 The teacher bought the books for the children.  
 "The teacher bought the books for the children."

The findings revealed that for both experiments, reading times for the critical region in Condition 1 (i.e., the condition where the functional head and its complement appeared in the same language) were significantly longer than reading times for the same region in Condition 2 (i.e., the condition where the language of the functional head and that of its complement were different). These findings

corroborated the results of corpus-analysis, replicating in reading times the preference patterns found in corpora frequencies: codeswitched constituents in which functional elements do *not* participate in the codeswitching process seem to be preferred over constituents in which functional elements undergo codeswitching. The convergence of data from these disparate sources lends support to the hypothesis that corpora and comprehension data are somehow interdependent. This, of course, raises the question of how the production and comprehension systems are interconnected. One possibility, discussed in Dussias (forthcoming), is to suggest that the production mechanism is constrained in a principled manner but the comprehension system is sensitive predominantly to information about statistical frequency. Although this seems a viable possibility, one limitation of the comprehension studies referred to above is that production data for the group of bilinguals that participated in them was not available. Hence, generalizations regarding the interaction between the comprehension and production systems in bilinguals are, at best, tentative. To examine this interaction in finer detail, the present study collected experimental production data from Spanish-English bilingual speakers who belonged to the same speech community as the bilinguals that participated in Dussias (1999). For comparison purposes, the two syntactic sites under investigation are, once again, Comp-IP and Det-NP.

## 2. *The present study*

The purpose of the present study is to investigate whether the comprehension preferences found in Dussias (1999) can be replicated in production data.

### 2.1 *Subjects*

Thirty-four subjects participated in this study. The subjects completed a language background survey designed to tap into several aspects of language proficiency and use by self-report. Although only approximately half of the subjects that participated in Dussias (1999) were available to participate in this study, the language survey revealed that all subjects belonged to the same speech community and had very similar language histories. The subjects were fluent Spanish-English bilinguals, had learned Spanish before the age of five, and English before seven years of age. Subjects reported using both languages in their daily lives with family and friends, and in a variety of contexts, including academic and non-academic. All subjects indicated that Spanish-English codeswitching was part of their daily linguistic behavior.

When asked about language dominance, 24% of the subjects indicated that they were equally proficient in both languages, 73% were English dominant and 3% were Spanish dominant.

## 2.2 *Materials and design*

Forty experimental items representing two conditions were used in this experiment. Condition 1 was designed to elicit the production of a determiner and Condition 2 to elicit the production of a complementizer. Items in each condition consisted of two phrases--one in English and one in Spanish--and each phrase was between three and seven words long. Given that determiners can appear in several positions within a sentence (i.e., subject noun phrase, object noun phrase, prepositional phrase, etc.), the stimuli for Condition 1 were constructed so that the target determiner always occupied the head position of an object noun phrase.

To ensure that order of presentation would not bias the results (i.e., if a sentence began in English, subjects would be more likely to produce an English determiner), order of presentation was counterbalanced so that for half of the items, the Spanish phrase appeared before the English phrase, and for the other half the English phrase appeared before the Spanish phrase. This yielded the resulting experimental items outlined in Table 1.

An additional set of 120 filler items was added to the experimental stimulus to serve as distracters. These sentences were similar in structure and length to the experimental sentences but required subjects to produce prepositions, question words, adverbs, quantifiers, and the copula 'be.'

## 2.3 *Procedure*

The instrument used for data collection was an elicited oral production task. Subjects were seated in front of a computer and, at the sign of a prompt, two phrases appeared on the screen, one below the other. Subjects were instructed to read both phrases aloud and to produce a complete sentence by combining the two phrases using only one word.<sup>3</sup> Subjects were told that the word could be either in English or Spanish, depending on their preference, and that there was no correct answer. For example, for the phrase '*La enfermera dijo \_\_\_\_\_ the patient didn't want to eat,*' subjects were expected to produce either '*La enfermera dijo que the patient didn't want to eat*' or '*La enfermera dijo that the patient didn't want to eat.*' For half of the sentences, the ending phrase was displayed on the screen prior to the beginning phrase, and for the other half the order was reversed.

Each time a sentence was presented to a subject, it was pseudo-randomly scrambled. This involved assigning an equal number of experimental and filler sentences to a number of different blocks, with the result that the items were

---

<sup>3</sup> Subjects were not given explicit instructions to use a complementizer or a determiner when constructing the sentences to ensure that the language samples produced were as natural as possible.

presented in a different order to each subject, yet the items in each condition were evenly distributed throughout the duration of the experiment. Finally, care was taken so that not more than two sentences representing the same condition were displayed sequentially on the computer screen and that six filler sentences were displayed between conditions.

The data was recorded in a sound-proof facility using a SONY digital recorder.

---

**Condition 1**

Spanish-to-English sample item

*El estudiante se olvidó*

The student forgot

“The student forgot

*history book in the library.*

history book in the library.

history book in the library.”

English-to-Spanish sample item

*The boy broke*

The boy broke

“The boy broke

*plato que estaba en la mesa.*

plate that was on the table.

plate that was on the table.”

---

**Condition 2**

Spanish-to-English sample item

*La enfermera dijo*

The nurse said

“The nurse said

*the patient didn't want to eat.*

the patient didn't want to eat.

the patient didn't want to eat.”

English-

to-Spanish sample item

*My sister thought*

My sister thought

“My sister thought

*su amiga estaba en la universidad.*

her friend was at the university.

her friend was at the university.”

---

Table 1: *Experimental design and item sample*

#### 2.4 Analysis and results

In order to determine the frequency with which subjects produced Spanish and English functional elements, the percentage of determiners and complementizers produced in both languages was calculated. The results for the determiner are given in Table 2 and for the complementizer in Table 3:

		Determiner	
		Spanish	English
Sentence	Spanish	56%	44%
beginning in	English	86%	14%

Table 2: Percentage of determiners produced in Spanish and English

		Complementizer	
		Spanish	English
Sentence	Spanish	84%	16%
beginning in	English	82%	18%

Table 3: Percentage of complementizers produced in Spanish and English

Table 2 shows that when the stimulus began in Spanish, subjects were more or less equally likely to produce a Spanish or an English determiner.

The results further reveal that when the stimulus began in English, subjects produced significantly more determiners in Spanish than in English (86% vs. 16%, respectively). These findings alone indicate that the language of the phrase preceding the determiner did not influence the subjects' choice of language at the time of production. If that had been the case, we would have expected to see a greater occurrence of, say, Spanish determiners when the stimulus began in Spanish.

Turning now to Table 3, we see that when the items began in Spanish, subjects produced the Spanish complementizer 'que' significantly more times than the English complementizer. Interestingly, the same preference was observed when the stimulus began in English, where subjects produced the Spanish complementizer 82% vs. 18% for the English complementizer.

What we see from the findings reported here is that during language production, some linguistic environments resist codeswitching more than others. Thus, for example, codeswitches where both Comp and IP appear in English preceded by a Spanish verb (e.g., La enfermera dijo that the patient didn't want to eat) are produced significantly less times than codeswitches where a Spanish Comp is preceded by a Spanish verb and followed by an English IP (e.g., La enfermera dijo que the patient didn't want to eat). This may very well reflect the strong bond that

exists between verbs and the complementizers they subcategorize for.<sup>4</sup> The 82% production preference observed for Spanish complementizers preceded by English verbs and followed by Spanish IP complements (as in ‘My sister thought que su amiga estaba en la universidad’) could simply indicate that the bond between a verb and a complementizer in English is somewhat weaker, thereby making codeswitching at this site more permissible. This intuition is partially supported by the fact that overt English complementizers are optional in the constructions under investigation.

In concluding, I return to the question of the nature of the relationship between the comprehension and production systems. One possibility, suggested above, is that linguistic, psycholinguistic and discourse principles may underlie the frequency patterns observable in spontaneous discourse for codeswitched utterances, and that the comprehension of these utterances takes place to a great degree independent of these principles. In this case, then, factors such as frequency of occurrence in the production may be said to be at least partially responsible for preference patterns observed in comprehension. The results presented here partly support this hypothesis. We saw, for example, that Spanish complementizers were produced significantly more times than English complementizers when preceded by Spanish verbs. This was precisely the pattern of preference observed in the comprehension data discussed above: subjects’ eye fixations were shorter when reading sentences with a Spanish complementizer preceded by a Spanish verb than when reading constructions with an English complementizer preceded by a Spanish verb. It is clear, however, that this account does not do an adequate job of explaining the results obtained for Spanish determiners. That is, although codeswitches between a Spanish determiner and an English noun are easier to understand (as measured by eye-movement fixations), there appears to be no clear preference for one codeswitch type over the other when it comes to production. A comprehensive account of the variables that may determine comprehension preferences in codeswitching remains to be investigated through future research. There are, however, a number of accounts that have been put forth to account for production preferences. For one, the source of the preference found in production data could be partly psycholinguistic, and partially rooted in the automaticity that characterizes function words. Because function words are less vulnerable to lexical access effects than content words, they are less likely to participate in codeswitching. In addition, function words are not accessed independently of syntactic information, are less often preceded by pauses than content words, are more predictable than content

---

<sup>4</sup> One may assume that this relationship manifests itself at different levels of linguistic representation. Thus, for example, one could expect to find phonological evidence (i.e., prosodic patterns) that is consistent with a view that the verb and its complementizer behave as one unit. This is the focus of Dussias (forthcoming).



words, and do not count as separate elements in motor planning units. Combined, these traits make function words less prone to undergo codeswitching (Muysken 1997).

An additional account put forth in Muysken (1997) deals with the low degree of '*insertability*' of functional elements vis-à-vis lexical elements into a syntactic frame. The argument goes something like this. A content word, say, a Spanish noun, would be easily insertable in an English frame because both the Spanish noun and the English noun it would replace instantiate the category N. In addition, since it is relatively easy to find a match between two content words in different languages (e.g., Spanish *casa* and English *house*), nouns are inserted quite effortlessly. Function words, on the other hand, are not easily definable because they are made up of highly specific features. Hence, finding a match between two function words is, presumably, more difficult. These factors result in a decrease in the involvement of function words in codeswitching.

A final account is provided by Myers-Scotton's Matrix Language Frame Model (1993, 1997; see also Myers-Scotton 1995). The model, which is psycholinguistically based, tries to account for codeswitching behavior by relating it to models of monolingual speech production. A key feature of the model is the distinction between Matrix Language and Embedded Language. The Matrix Language is defined as "the main language in codeswitching utterances...[It is the language that specifies] the morpheme order and supplies the syntactically relevant morphemes in constituents consisting of morphemes from both participating languages" (1993:3). The Embedded Language, on the other hand, refers to the other language that participates in codeswitching. Its role is less prominent than that of the Matrix Language, since, for the most part, it provides the content morphemes in codeswitched constituents. Drawing from the differential behavior of closed-class and open-class morphemes in monolingual speech production data, Myers-Scotton (1993, 1997) proposes a further distinction; that between 'system morphemes' and 'content morphemes'. The distinction is similar, but not identical, to the one made between closed-class items (i.e., determiner, complementizer, inflection, etc.) and open-class items (i.e., content words) in monolingual speech production studies. Myers-Scotton proposes that during codeswitched speech 'system morphemes' are expected to participate in codeswitching in a different way from 'content morphemes'. This is so because 'system morphemes' are selected following the morphosyntactic specifications set by the Matrix Language, whereas 'content morphemes' are supplied by the Embedded Language. The Matrix Language Frame Model provides an explanation for why, in codeswitching utterances involving Det + Noun ((1) through (3) above) and Comp + IP ((4)), the most frequent codeswitches are those where the determiner and the complementizer appear in Spanish and their

corresponding complements (i.e., Noun and IP) appear in English. In (1) above, for example, since the determiner is accessed according to the morphosyntactic specifications of Spanish (the Matrix Language), the Spanish determiner 'los' is accessed and inserted into the syntactic frame. English content morphemes that are congruent with the morphosyntactic specifications set by the Matrix Language are later inserted in the codeswitched constituent ('*holidays*' in this case), giving rise to a codeswitched constituent that has a Spanish system morpheme and an English content morpheme. This type of codeswitched constituency is, according to Myers-Scotton (1999, personal communication), the optimal type of constituent from a psycholinguistic perspective because, presumably, it is produced faster and with less effort than other types of codeswitches (i.e., whole noun phrase codeswitches).

In closing, we would like to note that we made the accidental discovery that certain types of determiners may be more prone to participating in codeswitching than others. In our data, if subjects produced a possessive determiner for sentences such as *María compró \_\_\_ vestido de novia en Nueva York* ("Mary bought \_\_\_ wedding dress in New York"), it invariably appeared in Spanish. However, when subjects produced an article, it appeared in Spanish at approximately the same rate as it appeared in English. Because the number of sentences of this type was small, no definite statements can be made. If substantiated with additional data, this finding may suggest that semantic factors are also at play when dealing with codeswitching at the level of speech production. At present, we are collecting additional data to corroborate this hypothesis.

## REFERENCES

- Cowper, E.A. 1992. *A Concise Introduction to Syntactic Theory*. Chicago, Ill.: University of Chicago Press.
- DiSciullo, Anne-Marie, Pieter Muysken & Rajendra Singh. 1986. "Government and code-mixing". *Journal of Linguistics* 22.1-24.
- Dussias, Paola E. 1997. "Sentence matching and the functional head constraint in Spanish/English codeswitching". *Spanish Applied Linguistics* 1.114-150.
- . 1999. "The function-word effect in Spanish-English codeswitching: A look at eye-movements". Paper presented at *the Second International Symposium on Bilingualism*, held in Newcastle upon Tyne, U.K., April 1999.
- . Forthcoming. "Psycholinguistic complexity in codeswitching". *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*.

- Halmari, Helena. 1997. *Government and codeswitching: Explaining American Finish*. Amsterdam & Philadelphia: John Benjamins.
- Klavans, Judith E. 1985. "The syntax of code-switching: Spanish and English". *Selected Papers from the 13<sup>th</sup> Linguistic Symposium on Romance Languages* ed. by L. D. King & C. A. Matey, 213-231. Amsterdam & Philadelphia: John Benjamins.
- Lance, David. 1975. "Spanish-English code-switching". *El lenguaje de los chicanos* ed. by E. Hernández-Chávez, A. Cohen & A. F. Beltrano, 138-153. Arlington, Va.: Center for Applied Linguistics.
- Lipski, John. M. 1985. "Linguistic aspects of Spanish-English language switching". *Series: Special Studies 25*. Arizona State University.
- Milian, Sarah. 1996. "Case assignment in Spanish/English Codeswitching". Paper presented at the *Symposium on Codeswitching*, Linguistic Society of America, annual meeting, **held in** , January 1996.
- Muysken, Pieter. 1997. "Code-mixing: Constraints and strategies". Paper presented at the *First International Symposium on Bilingualism*, held in Newcastle upon Tyne, U.K., April 1997.
- Myers-Scotton, Carol. 1993. *Duelling languages: grammatical structure in codeswitching*. Oxford: Oxford University Press.
- . 1995. "A lexically based model of code-switching". *One speaker, two languages: Cross-disciplinary perspectives on code-switching* ed. by Lesley Milroy & Pieter Muysken, 233-256. Cambridge: Cambridge University Press.
- . 1997. *Duelling languages: grammatical structure in codeswitching* (2<sup>nd</sup> ed.). Oxford: Oxford University Press.
- & Janice L. Jake. 1997. "Codeswitching as evidence for two types of functional elements". Paper presented at the *First International Symposium on Bilingualism*, held in Newcastle upon Tyne, U. K., April 1997.
- Park, J., Rudolf Troike & M. Park. 1993. "Constraints in Korean-English Code-Switching: A Preliminary Study". *Journal of the Applied Linguistic Association of Korea* 6.115-133.
- 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. Also in *Spanish in the United States, sociolinguistics aspects* ed. by J. Amastae & L. Elias-Olivares. 1982, 230-263. Cambridge: Cambridge University Press.
- Sankoff, David & Shana Poplack. 1981. "A formal grammar of code-switching". *Papers in Linguistics* 14.3-43.

Treffers-Daller, Jenine. 1995. "Code-switching between a case assigner and its complement: The case of Turkish-German prepositional phrases". *Summer School codeswitching and language contact, 14-17 Sept. 1994*. Ljouwert/Leeuwarden, The Netherlands: Fryske Akademy.