

Simulating Spanish-English Code-Switching: *El Modelo Está* Generating Code-Switches

Chara Tsoukala
Centre for Language Studies,
Radboud University
c.tsoukala@let.ru.nl

Stefan L. Frank
Centre for Language Studies,
Radboud University
s.frank@let.ru.nl

Antal van den Bosch
KNAW Meertens Institute;
Centre for Language Studies,
Radboud University
a.vandenbosch@let.ru.nl

Jorge Valdés Kroff
Department of Spanish and Portuguese Studies,
University of Florida
jvaldeskroff@ufl.edu

Mirjam Broersma
Centre for Language Studies,
Radboud University
m.broersma@let.ru.nl

Abstract

Multilingual speakers are able to switch from one language to the other (“code-switch”) between or within sentences. Because the underlying cognitive mechanisms are not well understood, in this study we use computational cognitive modeling to shed light on the process of code-switching. We employed the Bilingual Dual-path model, a Recurrent Neural Network of bilingual sentence production (Tsoukala et al., 2017) and simulated sentence production in simultaneous Spanish-English bilinguals. Our first goal was to investigate whether the model would code-switch without being exposed to code-switched training input. The model indeed produced code-switches even without any exposure to such input and the patterns of code-switches are in line with earlier linguistic work (Poplack, 1980). The second goal of this study was to investigate an auxiliary phrase asymmetry that exists in Spanish-English code-switched production. Using this cognitive model, we examined a possible cause for this asymmetry. To our knowledge, this is the first computational cognitive model that aims to simulate code-switched sentence production.

1 Introduction

People who speak several languages are able to switch from one language to the other, between or within sentences, a process called code-switching. Code-switching has been studied for decades by theoretical linguists and sociolinguists (e.g., Poplack 1980; Muysken 2000) and more recently by psycholinguists (e.g., Bullock and Toribio 2009). In the past few years it has started being studied with a computational methodology, and it has garnered attention among the natu-

ral language processing (NLP) research community. Several NLP applications have emerged, e.g., to detect code-switches (Solorio and Liu, 2008; Guzmán et al., 2017), or to automatically recognize code-switched speech (Yılmaz et al., 2016; Gonen and Goldberg, 2018). Moreover, there are a small number of cognitive computational models relevant to code-switching: Filippi et al. (2014) developed a model of code-switched word production and Janciauskas and Chang (2018), while simulating age of acquisition effects on native Korean speakers of English, reported that the models that had been exposed to English later produced code-switches, i.e., occasionally used Korean words in their predominantly English production.

The underlying mechanisms of code-switching, however, are still not well understood. Therefore, we suggest using computational cognitive modeling to simulate code-switching behavior in multilinguals with the goal of gaining more insight into the process of code-switching. In this work, we have employed a model of bilingual sentence production (Tsoukala et al., 2017) and tested whether it can produce spontaneous code-switches without being trained on code-switched sentences (Experiment 1, Section 3). We wanted to test whether code-switching can be (partially) attributed to internal factors and explained by the distributions of the two languages involved, or whether it is strictly a community-based practice that can only be explained by exposure to code-switches. To test the former, we hypothesized that a model that receives training input in two languages but no code-switched sentences, will nevertheless be able to produce code-switched sentences by combining patterns from the two languages it has been

trained on. We then employed this model to investigate a specific production phenomenon that occurs in Spanish-English code-switching (Experiment 2). As explained in Section 4, we wanted to test whether the phenomenon of interest is caused by the distributional properties of the two languages. This is something that can be explicitly tested with this model because it is not trained on code-switched input.

To our knowledge, this is the first computational cognitive model that aims to simulate code-switched sentence production.

2 Model

To simulate code-switched sentence production, we first needed to simulate bilingual production. For that purpose, we employed the Bilingual Dual-path model (Tsoukala et al., 2017) and trained it to simulate simultaneous Spanish-English bilinguals, i.e., speakers who acquired both Spanish and English from infancy.

The Bilingual Dual-path model is a modified version of Dual-path (Chang, 2002). We chose to work with, and extend, the Dual-path model because it is one of the most successful and empirically validated cognitive models of sentence production. It has been used to explain a wide range of phenomena in various languages; for an overview see Frank et al. (in press).

2.1 Bilingual Dual-path Model

The Bilingual Dual-path model (Figure 1) is a Recurrent Neural Network (RNN) based on the Simple Recurrent Network (SRN; Elman 1990) architecture. It learns to convert a message into a sentence by predicting the sentence word by word. Dual-path got its name because of its two pathways that influence the production of each word: i) the meaning, or semantic, system that learns to map words onto concepts (and their realization, see below and Section 2.2.1), thematic roles, event semantics and the intended language (“target language”), and ii) the sequencing, or syntactic, system that is an SRN that learns to abstract syntactic patterns. Both paths influence the next word prediction (the “output” layer).

To express a new message (see Section 2.2.4 for examples of messages), the following items are fixed and influence the production of the first word: the to-be-expressed semantic roles have fixed connections with their concepts and realiza-

tions, and the relevant “event semantics” and “target language” units are activated. Additionally, the hidden layer’s context units are reset to a default value (0.5 in our simulations).

The output word is determined as the word with the highest activation in the output layer. Once an output word has been produced, it is fed back as input (to the “input” layer). During the training phase, the target word is given as input instead of the (potentially different) output word.

The sequencing system is a regular SRN that has one recurrent hidden layer (of 110 units in our simulations) and two 70-unit “compress” layers that are placed between the input word and the hidden layer, and between the hidden layer and the output word.

The meaning system learns to map the input word onto a concept and, whenever relevant, the realization of that concept (PRON for pronoun, INDEF for an indefinite article and DEF for definite articles; see Section 2.2.1 (“Message”) for concrete examples).

A difference between this architecture and other RNNs is that whenever a new message needs to be expressed, the network receives fixed connections between concepts and roles; this allows their separation (instead of having a single unit for, e.g., ‘AGENT-WOMAN’) and, in turn, enables the model to generalize and to produce words in novel roles. For instance, if the concept ‘WOMAN’ has only been seen as an AGENT in the training set, it can still be correctly expressed in novel roles (PATIENT, RECIPIENT) during the test phase (Chang, 2002).

All layers use the *tanh* activation function, except the output and predicted role layers that use *softmax*. The model is built in Python and can be found at <https://github.com/xtsoukala/CMCL19>.

2.2 Input Languages

In order to simulate Spanish-English bilingual sentence production, we generated input with relevant properties of the two languages. The sentences (and their messages, see Section 2.2.1) are generated before the training starts, and they are based on the allowed structures (Section 2.2.2). For each part of speech (POS) a randomly selected lexical item (from that POS and target language) is sampled from the lexicon (Section 2.2.3). The advantage of using artificial (miniature) languages is

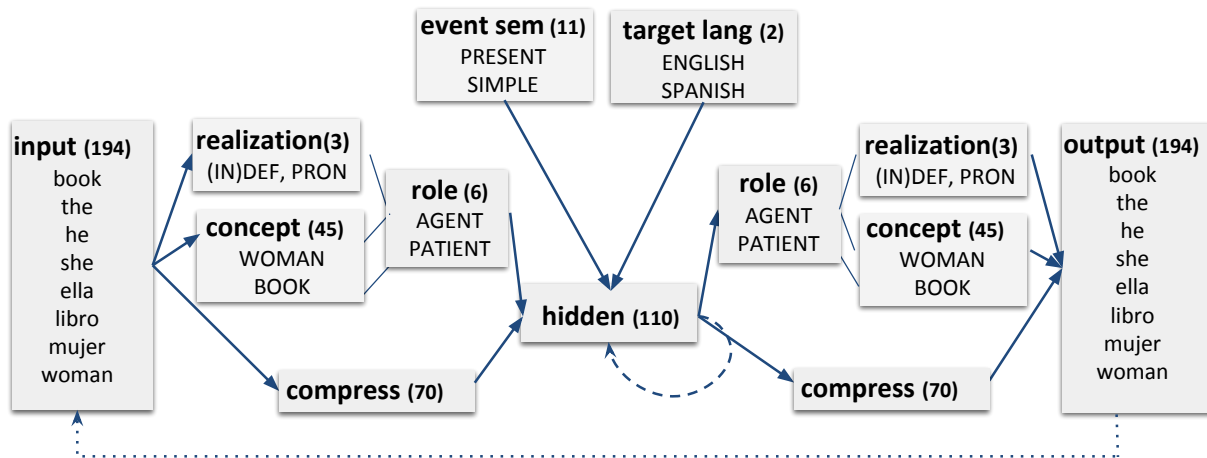


Figure 1: Bilingual Dual-path, the model used in these simulations, is a next-word prediction model that converts messages into sentences. It is based on an SRN (the lower path, via the ‘compress’ layers) that is augmented with a semantic stream (upper path) that contains information about concepts, thematic roles, event semantics, and the target language.

that we can manipulate the frequency and grammar of the input and isolate (and thereby study) the phenomenon of interest.

2.2.1 Message

The model is trained using generated sentences (as described above) paired with their message that consists of semantics and their realization, event-semantics, and target language, which will be explained in turn below.

In these simulations, the semantics contains information regarding 45 unique concepts and 6 thematic roles: AGENT, AGENT-MODIFIER, PATIENT, ACTION-LINKING, RECIPIENT, and ATTRIBUTE.

ACTION-LINKING is a combined thematic role that can be used for all verb types: action (e.g., ‘throws’), linking (‘is’) and possession (‘has’). ATTRIBUTE is an attribute expressed with a linking verb concept (‘BE’). AGENT and RECIPIENT can be expressed only with animate nouns.

A concept (e.g., WOMAN for the English word ‘woman’ or Spanish word ‘mujer’) is assigned to each thematic role (during the sentence generation process) along with a realization attribute (PRON for pronoun, DEF for definite article, and INDEF for indefinite article) according to the meaning that needs to be expressed. For instance, in the sentence “the woman runs” the message would include “AGENT=WOMAN, DEF”, whereas “a woman” would be encoded as “AGENT=WOMAN, INDEF”, and “she” as “AGENT=WOMAN, PRON”.

Furthermore, the message contains event semantic information (denoted as EVENT-SEM), which gives information regarding the tense (PRESENT, PAST) and aspect (SIMPLE, PERFECT or PROGRESSIVE). The EVENT-SEM layer also contains information regarding the roles needed for that particular message; the model needs to keep track of the roles expressed and make sure that if, e.g., the role of RECIPIENT is activated then the recipient has also been expressed.

Additionally, the message contains information about the target language so that the model knows whether it is learning to produce an English or Spanish sentence.

2.2.2 Structures

The allowed structures for both languages are the following in our simulations:

- SV: Subject - Verb, e.g., “a happy dog runs”; “un perro feliz corre”
- SVO: Subject - Verb - Object, e.g., “the boy is carrying a book”; “el niño está llevando un libro”
- SVDOPP: Subject - Verb - Direct Object - Prepositional Phrase, e.g., “she shows a book to the girl”; “ella muestra un libro a la niña”
- SVIDODO: Subject - Verb - Indirect Object - Direct Object: e.g., “she shows the girl a book” (Structure occurs only in English)

- SVPPDO: Subject - Verb - Prepositional Phrase - Direct Object: e.g., “ella muestra a la niña un libro” (“she shows to a girl a book”). Structure only used in Spanish.

The roles can be expressed using either a Noun Phrase (NP) with definite (DEF) or indefinite (INDEF) article (e.g., ‘the woman’, ‘a woman’). Additionally, AGENT can be expressed with a pronoun (PRON, e.g. ‘she’). NPs optionally contain a modifier (an adjective, e.g., ‘a tall woman’).

The verbs are either intransitive (e.g., ‘sleep’), transitive (‘carry’), double transitive (‘show’), linking (‘is’) or possession verb (‘has’). The tense is present or past and the aspect is simple, progressive, or perfect. Only the simple past was used whereas the present tense is used with all three aspects:

- simple present: “the man cooks”; “el hombre cocina”
- present progressive: “the man is cooking”; “el hombre está cocinando”
- present perfect: “the man has cooked”; “el hombre ha cocinado”

2.2.3 Bilingual Lexicon

The lexicon consists of 194 words (Table 1): 88 English words, 105 Spanish words, and the shared period (‘.’) that marks the end of the sentence. The Spanish lexicon is bigger because this language is gendered: for instance, ‘tired’ is either ‘cansado’, if it modifies a masculine noun, or ‘cansada’ for a feminine noun. Syntactic category information (such as ‘adjective’, ‘participle’) is not given explicitly; the model learns it through the syntactic path during training.

2.2.4 Input Examples

To illustrate the input, here is an example of the message (excluding the target language):

AGENT=WOMAN, INDEF
 AGENT-MOD=TALL
 ACTION-LINKING=GIVE
 PATIENT=BOOK, DEF
 RECIPIENT=GIRL, DEF
 EVENT-SEM=SIMPLE,PRESENT,AGENT,
 AGENT-MOD, PATIENT, RECIPIENT

POS	<i>n</i>	Examples
Verbs	64	
auxiliary	4	is, has, <i>está, ha</i>
intransitive	32	walked, swims, <i>nada</i>
transitive	12	carries, push, <i>lleva</i>
double transitive	12	gives, throws, <i>da</i>
possession	4	has, had, <i>tiene, tenía</i>
linking ¹	4	is, was, <i>está, estaba</i>
Participles ²	57	eating, eaten, <i>comido</i>
Nouns	46	
animate	10	uncle, aunt, <i>tío, tía</i>
inanimate	36	pen, book, <i>libro</i>
Adjectives	22	busy, <i>ocupado</i>
Determiners	6	a, the, <i>un, una, el, la</i>
Prepositions	2	to, <i>a</i>
Pronouns	4	he, she, <i>él, ella</i>

¹ Three of these overlap with the auxiliary verbs.

² Nine of these have the same form as a verb; e.g., ‘walked’ is either a perfect participle or a verb.

Table 1: POS in bilingual lexicon (Spanish in italics)

This message would be expressed linguistically in the following manner in English and Spanish:

- a tall woman gives the girl a book .
- una mujer alta da a la niña un libro . (word-by-word translation: “a woman tall gives to the girl a book”)

If the aspect was PROGRESSIVE instead of SIMPLE, on the other hand, the corresponding sentences would be “a tall woman is giving the girl a book”; “una mujer alta está dando a la niña un libro”.

The linking verb messages were encoded in the following manner:

AGENT=WOMAN, DEF
 ACTION-LINKING=BE
 ATTRIBUTE=TIED
 EVENT-SEM=SIMPLE,PRESENT,
 AGENT, ATTRIBUTE

and expressed as “the woman is tired”; “la mujer está cansada”.

2.3 Training

The model was trained on a total of 3040 randomly generated sentence-message pairs in English and

Spanish (training set; 50% [1520 pairs] per language). Recall that no code-switched sentences were given as input.

We ran 60 simulations using different input and different random initial weights per simulation, as the input and the weights are the only non-deterministic parts of the model. The models were trained for 30 epochs, where 1 epoch corresponds to a full iteration of the training set (3040 sentences). At the beginning of each epoch, the training set was shuffled.

The “realization–role” and “role–realization” connection weights were set to 10, and the “concept–role” and “role–concept” to 30. The initial learning rate was 0.10 and linearly decreased over 10 epochs until it reached 0.02; the momentum was set to 0.9. None of the hyper-parameters was optimized for the task, and they do not play a crucial role in the results. We selected the values from Tsoukala et al. (2017) and increased the “concept–role” connections because this resulted in slightly better performance (the current experiments use more concepts).

2.4 Evaluation and Performance Threshold

The correctness of a sentence is determined by whether the correct (and complete) semantic meaning has been expressed in a grammatical sentence but not necessarily in the target syntactic structure. For instance, if the target sentence is “a sad grandfather is showing the book to a girl” and the produced sentence is “a sad grandfather is showing the *pen* to a girl” it is counted as incorrect, whereas if the produced sentence is “a sad grandfather is showing *the girl the book*” it is counted as correct even though it was expressed with a different syntactic structure than the target one. If it is expressed with a different aspect (e.g., perfect instead of progressive) or realization (e.g., pronoun instead of an NP with an indefinite article) it is also marked as incorrect. If the sentence contains code-switches it is marked as correct as long as it expresses the correct meaning, is expressed in one of the allowed structures (Section 2.2.2) and the POS sequence of each phrase (NP, Verb Phrase [VP], Prepositional Phrase [PP]) is valid in either language.

For all the experiments, we excluded from the analysis simulations that did not learn to produce at least 75% of the messages correctly according to the criteria above.

3 Experiment 1: Code-Switching

In this study, we investigate whether the Bilingual Dual-path would produce code-switched sentences if trained on Spanish and English (but not code-switched) sentences. We investigate the occurrence of different patterns of code-switching that have been observed in the language use of human bilingual speakers.

3.1 Background

Muysken (2000) proposed the following typology of code-switching:

1. Insertional switching

Insertions of single words/fixed expressions:

- lexical (e.g., noun): “I read a *libro*” (I read a book)
- fixed expressions/ interjections/ idiomatic expressions: “*Oh my god*, estamos sin palabras” (we are speechless)

2. Alternational switching

Alternation between the two languages, involving multi-word sequences, either between or within sentences:

- Inter-sentential switching: “I heard you had an accident. *¿Qué pasó?*” (What happened?)
- Intra-sentential switching: “I had a hard time finding *tu casa esta mañana*.” (your house this morning)

3. Congruent lexicalization

In cases where the languages share syntactic structures and are highly cognate, it is possible to use the shared syntax and insert lexical items from either language, thus seemingly switching back and forth: e.g., “*Bueno*, in other words, *el flight que sale de Chicago* around three o’clock” (‘Fine, ... the flight which leaves from ...’) (Pfaff, 1979)

3.1.1 Code-Switching by Syntactic Category

In a seminal study, Poplack (1980) observed the Puerto-Rican community in the US. She found that balanced bilinguals produced mostly complex code-switches, such as intra-sentential ones, and few insertions. Switches at the NP were more frequent than switches at the VP and PP, and noun insertions were the most frequent lexical insertion whereas determiner insertions occurred rarely.

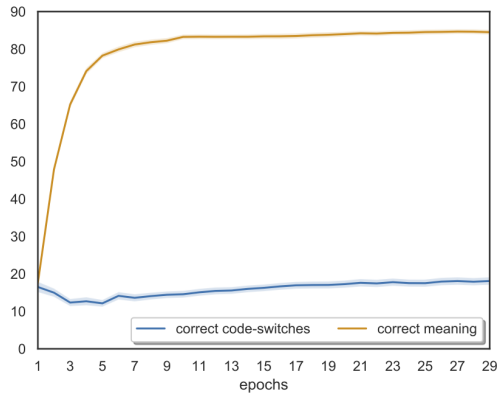


Figure 2: Percentage of correctly produced sentences and of code-switches among those sentences. The shaded area shows the Standard Error of the Mean (SEM) computed over 56 simulations

3.2 Method

To simulate code-switching, we trained the model as described in Section 2.3 and tested it on 760 unseen sentences (test set) that were randomly generated in the same manner as the training set.

During the test (“production”) phase we manipulated the model’s language control by activating a target language only at the beginning, before the production of the first word, so as to indicate the conversational setting (intended language). After the first word had been produced, we activated both target language nodes, thus allowing the model to produce the sentence in either language or to code-switch.

We excluded from the analysis four models that did not pass the 75% performance threshold (as explained in Section 2.4). The reported results are from the remaining 56 simulations.

3.3 Results

As hypothesized, the model produced code-switches even though it had not been exposed to code-switched input. The model code-switched in 18.09% of the correctly produced sentences (at the last epoch, see Figure 2).

3.3.1 Typology of Code-Switching in the Model’s Output

Figure 3 shows the insertions per POS and the alternational code-switched types (per POS at which the first language switch occurred) that were produced by the model at the end of the training (30th epoch). The model produced alternational

switches more frequently than insertional switches (13.57% vs 4.52%).

3.3.2 Examples of Code-Switched Sentences

Insertional code-switches of different syntactic categories are illustrated below:

- Noun insertion:

Target: un anfitrión feliz ha pateado un bolígrafo . (English: a happy host has kicked a pen)

Output: un anfitrión feliz ha pateado un *pen* .

- Verb insertion:

Target: un camarero llevó la llave . (English: a waiter carried the key)

Output: un camarero *carried* la llave .

- Determiner insertion:

Target: he is showing the book to the father .

Output: he is showing *el* book to the father .

- Adjective insertion:

Target: a man is sad . Output: a man is *triste* .

Examples of alternational switches are provided below:

- Alternation at the determiner (Noun Phrase):

Target: the uncle has shown a father the toy .

Output: the uncle has shown *un padre* the toy .

- Alternation at the noun:

Target: the short boy shows a brother a book .

Output: the short boy shows a *libro a un hermano* .

- Alternation at the preposition (Prepositional Phrase):

Target: the tall waiter has given a brother a book .

Output: the tall waiter has given *a un hermano un libro* .

- Alternation at the auxiliary verb (Auxiliary Phrase):

Target: the short waiter is showing a dog a toy .

Output: the short waiter *está mostrando a un perro un juguete* .

Note that in the third example (Prepositional Phrase) the model inserted a preposition when switching, thus adhering to Spanish grammar: The

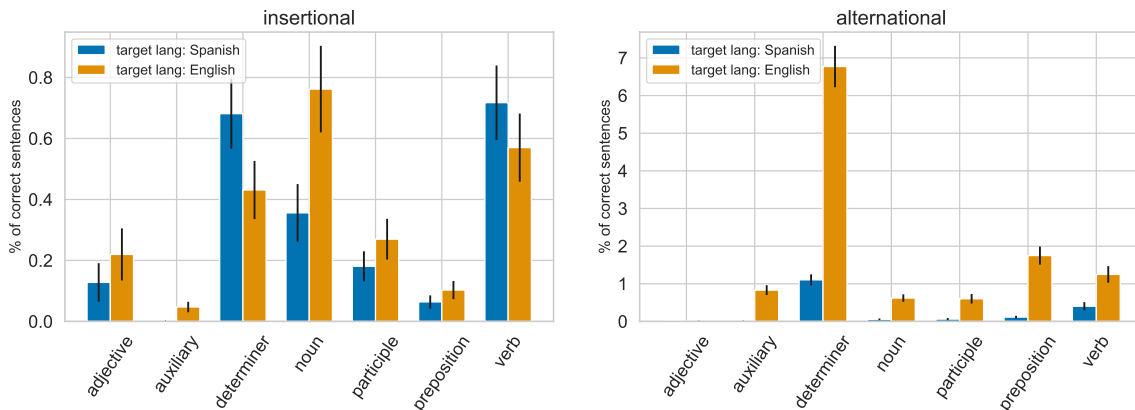


Figure 3: Types of insertional switching (left) and alternational switching (right). For alternational switches, the POS indicates the first point of switch. All values designate the percentage of correctly produced sentences. The error bars show the SEM computed over 56 simulations.

double dative does not exist with the double noun phrase form in Spanish. This cross-linguistic difference is even more relevant in the fourth example (Auxiliary Phrase switch) because the verb is in Spanish and the sentence would have been entirely ungrammatical if the model had not inserted a preposition (“a un perro”).

3.4 Discussion

The model produced spontaneous code-switches through the manipulation of the target language, without being exposed to code-switched input. This supports the hypothesis that code-switches can occur due to internal and distributional factors, and not only because of exposure to code-switching.

Simulating a balanced bilingual speaker, the model produced mostly alternational switches as opposed to insertional ones. This is in line with Poplack’s (1980) observation. Furthermore, alternations at the NP (alternational switch at the determiner) were more likely than alternations at the VP (alternational switch at the verb) or PP (alternational switch at the preposition), which is also in line with the patterns observed by Poplack. However, the model also produced code-switching patterns that are not attested in humans. For instance, the model inserted determiners (1.11% of the correctly produced sentences), especially English determiners in an otherwise Spanish sentence (0.68% of correctly produced sentences). We hypothesize that the model has this preference because determiners in English are not gendered. This means both that the model does not need to select a gendered article and that it prefers to use

the English determiner which has twice the frequency of the Spanish ones (as, e.g., ‘the’ is the translation of both ‘el’ and ‘la’ that are the Spanish definite determiners for masculine and feminine nouns respectively).

In bilingual environments where both languages are used, bilingual speakers start with an intended language that is defined by the conversational environment, but they are capable of communicating using either of their languages, or by code-switching (Grosjean, 2001). The top-down language control manipulation in the model (i.e., activating both target languages) is analogous to manipulating the conversational setting in which a speaker is interacting. Spontaneous code-switches occur when there is no target language preference. We only activate a target language right before the production of the first word so as to set the conversational environment.

4 Experiment 2: Auxiliary Phrase Asymmetry

Our second experiment applies the model to a specific code-switching phenomenon: a production asymmetry that has been observed among Spanish-English communities in the US.

4.1 Background

Spanish-English bilinguals are moderately likely to code-switch in the progressive structure between the Spanish auxiliary “estar” (“to be”) and the participle. For instance:

1. *Las personas están* protesting (The people are protesting)

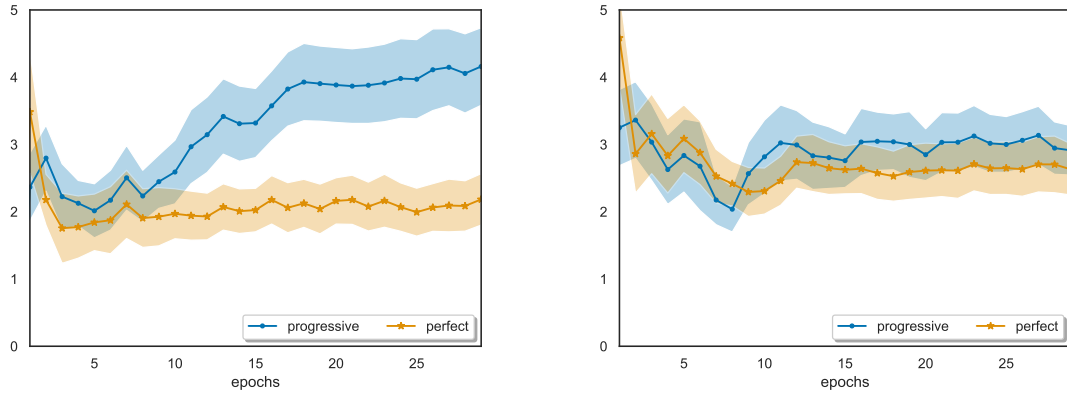


Figure 4: Percentage of Spanish-to-English participle switches for the correctly produced sentences per aspect in the “haber-model” (left) and the “tener-model” (right). Shaded areas show the SEM computed over 47 simulations.

is valid, whereas a switch at the perfect structure is rarely produced between the Spanish auxiliary “haber” (“to have”) and the participle:

2. * *Las personas han* protested (The people have protested)

Furthermore, a switch at the auxiliary is likely for both structures: “*Las personas are* protesting”, “*Las personas have* protested”.

This phenomenon is known as the “auxiliary phrase asymmetry” (Guzzardo Tamargo et al., 2016; Poplack, 1980; Pfaff, 1979), and it has been confirmed both in production through corpus analysis and in comprehension through an eye-tracking-while-reading study (Guzzardo Tamargo et al.).

According to the “grammaticalization account”, the source of this asymmetry is that “estar” has more semantic weight and is syntactically more independent as it also functions as a linking verb (e.g., “el enfermero está cansado”; “the nurse is tired”), whereas “haber” is highly grammaticalized as it is almost exclusively used as an auxiliary. The verb of possession in Spanish is “tener” (“el enfermero tiene un libro”; “the nurse has a book”), and “haber” is only used as an auxiliary verb or in archaic formulations (Guzzardo Tamargo et al., 2016). An alternative hypothesis suggested, but not attested, by Guzzardo Tamargo et al. is that the asymmetry emerges from community-supported practice (“exposure-based account”), i.e., speakers must be exposed to the production asymmetry.

4.2 Method

To investigate if the model provides support for the grammaticalization account, we first tested

whether the asymmetry would emerge in the model described in Section 2.3; this would imply that the asymmetry can emerge even only from the distributional patterns of the two languages and that the exposure-based account is not necessary to explain the phenomenon, as the asymmetry is not present in the input. Second, we took advantage of the fact that the model input is generated and can therefore be manipulated, and we tested explicitly the grammaticalization hypothesis by replacing all instances of “haber” with “tener”, the Spanish main verb “to have”, thereby adding semantic weight to the Spanish auxiliary verb.

Specifically, for the first research question we employed again the model described in Section 2.3 (“haber-model”), and we tested it on 1000 novel messages: 500 progressive sentences (e.g., “the boy is kicking a ball”) and 500 that were the perfect-tense equivalent of those sentences (“the boy has kicked a ball”). As in the previous experiment (Section 3), we activated both languages after the first word of each sentence had been produced.

For the second model (“tener-model”), we replaced all the instances of “haber” with “tener” in the training set (e.g., “el niño ha comido”; “the boy has eaten” became “el niño tiene comido”). We kept everything else the same (1000 test messages, initialized weights, lexicon size even though “haber” was no longer used, and all the layer sizes), and we ran 60 simulations using the modified input. Because “tener” is also used as an independent main verb with semantic content, we hypothesized that this model will not show the asymmetry.

In order to have a fair comparison between the two models, we only analyzed the simulations that had successfully learned to produce at least 75% of the sentences for both models.¹ Five simulations failed in the “haber-model” model and eight in the “tener-model”, thus leading to a total of 47 simulations to be analyzed.

4.3 Results

Even without any manipulations, the “haber-model” showed a strong preference for progressive participle switches: 4.16% vs 2.18% for the perfect participle switches. Figure 4 (left) shows the average percentage of Spanish-to-English participle switches over 47 simulations per aspect (progressive and perfect).

Figure 5 shows the percentage of code-switches at the auxiliary verb and participle for the progressive and perfect structure, after 30 training epochs. In the progressive aspect, the simulations did not show a preference for a switch at the auxiliary; it is equally likely compared to a switch at the participle, thus reflecting prior experimental and corpus-based results (Guzzardo Tamargo et al., 2016). In the perfect structures, on the other hand, a switch at the participle is much less likely than a switch at the auxiliary position.

When tested on the same 1000 messages, the “tener-model” (that substituted the original Spanish auxiliary verb from the “haber-model” for one with more semantic weight) did not show a preference for progressive participle switches (2.91% vs 2.63%; Figure 4, right).

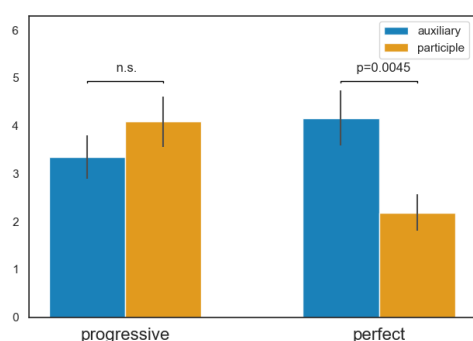


Figure 5: Percentages of code-switches at the auxiliary and participle for the progressive and perfect aspect, after 30 epochs. Error bars show the SEM computed over 47 simulations.

¹The “sameness” of the simulations is judged by the simulation number which indicates the seed for the initialized weights and the generated input.

4.4 Discussion

We tested whether the auxiliary phrase asymmetry in Spanish-English code-switching could be derived from the properties of the two languages. The “haber-model” simulated the attested asymmetry and the “tener-model” tested whether the cause could be attributed to the Spanish auxiliary “haber” that only has a limited, dependent syntactic function (i.e., is more grammaticalized) and is not used as frequently as the English equivalent (“have”). The two simulations confirm that the grammaticalization account could be responsible for the asymmetry.

5 Conclusion

We have presented a novel method to test hypotheses in code-switched sentence production. This computational cognitive model can easily be modified to simulate code-switched production of a different language pair. Additionally, the generated input allows for manipulations that help test other hypotheses about code-switching, for instance the idea that cognates can trigger code-switched speech (Clyne, 1980).

Acknowledgments

The work presented here was funded by the Netherlands Organisation for Scientific Research (NWO) Gravitation Grant 024.001.006 to the Language in Interaction Consortium.

References

- Barbara E. Bullock and Almeida Jacqueline Toribio. 2009. Trying to hit a moving target: On the sociophonetics of code-switching. *Multidisciplinary Approaches to Code Switching*, 41:189–206.
- Franklin Chang. 2002. Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26:609–651.
- Michael G. Clyne. 1980. Triggering and language processing. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 34(4):400.
- Jeffrey L. Elman. 1990. Finding structure in time. *Cognitive Science*, 14(2):179–211.
- Roberto Filippi, Themis Karaminis, and Michael S.C. Thomas. 2014. Language switching in bilingual production: Empirical data and computational modelling. *Bilingualism: Language and Cognition*, 17(2):294–315.

- Stefan L. Frank, Padraic Monaghan, and Chara Tsoukala. in press. [Neural network models of language acquisition and processing](#). In Peter Hagoort, editor, *Human Language: from Genes and Brains to Behavior*. The MIT Press.
- Hila Gonen and Yoav Goldberg. 2018. [Language modeling for code-switching: Evaluation, integration of monolingual data, and discriminative training](#). *arXiv preprint arXiv:1810.11895*.
- François Grosjean. 2001. [The bilingual's language modes](#). In Janet L. Nicol, editor, *One mind, two languages: Bilingual language processing*, pages 1–22. Oxford: Blackwell.
- Gualberto A Guzmán, Joseph Ricard, Jacqueline Serigos, Barbara E Bullock, and Almeida Jacqueline Toribio. 2017. [Metrics for modeling code-switching across corpora](#). In *INTERSPEECH*, pages 67–71.
- Rosa E. Guzzardo Tamargo, Jorge R. Valdés Kroff, and Paola E. Dussias. 2016. [Examining the relationship between comprehension and production processes in code-switched language](#). *Journal of Memory and Language*, 89:138–161.
- Marius Janciauskas and Franklin Chang. 2018. [Input and age-dependent variation in second language learning: A connectionist account](#). *Cognitive Science*, 42:519–554.
- Pieter Muysken. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge University Press.
- Carol W. Pfaff. 1979. [Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English](#). *Language*, pages 291–318.
- Shana Poplack. 1980. [Sometimes I'll start a sentence in Spanish y termino en Español: toward a typology of code-switching](#). *Linguistics*, 18(7-8):581–618.
- Thamar Solorio and Yang Liu. 2008. [Learning to predict code-switching points](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 973–981. Association for Computational Linguistics.
- Chara Tsoukala, Stefan L. Frank, and Mirjam Broersma. 2017. [“He’s pregnant”: simulating the confusing case of gender pronoun errors in L2](#). *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 3392–3397.
- Emre Yılmaz, Henk van den Heuvel, and David van Leeuwen. 2016. [Investigating bilingual deep neural networks for automatic recognition of code-switching Frisian speech](#). *Procedia Computer Science*, 81:159–166.