The Distance Hypothesis for Bilingual Code-switching Tested on German/English and Chinese/English Data

Eva Maria Duran Eppler
University of Roehampton, London

Lin Wang
College of Foreign Languages, Zhejiang Gongshang University, Hangzhou, China

Abstract

That closely related words tend to be close together in the sentence is a well-established principle of monolingual language comprehension and production. This paper suggests that this is different in bilingual language use. It proposes that long dependency distances between syntactically related units facilitate bilingual code-switching (Distance Hypothesis DH). Code-switching (CS) is the linguistic behaviour of producing or comprehending language that is composed from lexical items and grammatical structures from two (or more) languages. Dependency distance is the number of words between a head and a dependent. We test the DH on a 9,023 word German/English, and a 19,766 word Chinese/English corpus. In both corpora mixed dependencies present longer dependency distances than monolingual ones, which supports the DH. Selected major dependency types (subject, object, adjunct) also have longer dependency distances when the head word and its dependent are from different languages. We discuss how processing motivations behind the DH make it a potentially viable motivator for bilingual code-switching and - more generally - for contact-induced language change.

1. Introduction

Corpus linguistic, computational linguistic and experimental language research has produced a considerable body of evidence over the past thirty years that there is a preference for linguistically related words to be close together in monolingual sentences. Hudson (1995), Gibson (1998, 2000), Liu (2008) and others have worked on this from the comprehension side; Hawkins (1994, 2004), Temperley (2008) and collaborators have addressed the production side.

Most of this research captures the notion of "linguistically related" and "close together" with the concept of dependency distance/length. Dependencies are asymmetric syntactic relations between two words, a head/governor and a dependent. The head of each dependency is then the dependent of another word (unless it is the head/root of the sentence ↓), forming a recursive structure that connects all the words of the sentence, as in figure 1. Arrows point from the head to the dependent.

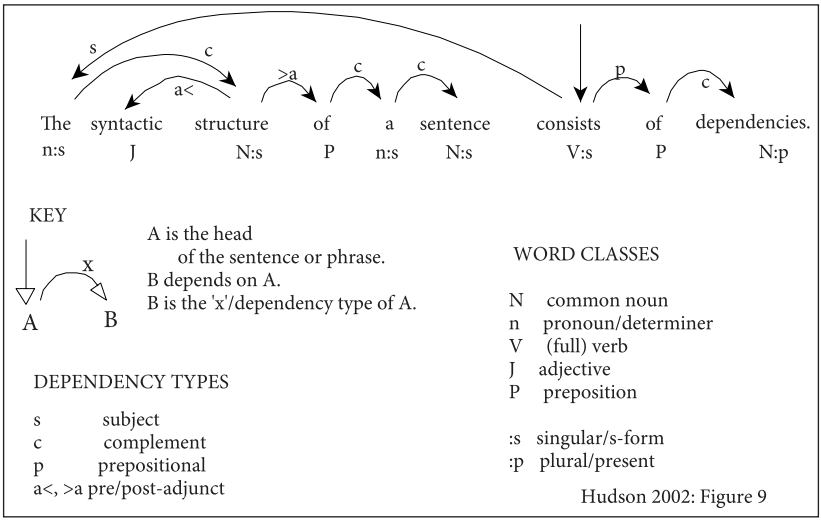

Figure 1. Dependency structure of an English sentence

Figure 1 illustrates that dependencies are (a) of a certain type, (b) directed, and (c) have a length (the distance between the head and the dependent).

a. Dependencies can be semantic, morphological and/or syntactic. In this paper we are only looking at syntactic dependencies; the arrows are therefore labelled for grammatical functions, e.g. subject, adjunct etc.
b. Every dependency arrow points from the head to the dependent, i.e. dependencies have direction. In a head-initial dependency the head precedes the (post-)dependent; in a head-final dependency, the (pre-) dependent precedes the head in the linear order of the sentence. Many languages have a dominant dependency direction: Arabic is predominantly head initial, Turkish head final; other languages, including English, German and Chinese, are more or less mixed.
c. Every arrow spans a specific number of words (unless it indicates the root of the sentence ↓). The linear distance between a head and its dependent, measured in terms of intervening words, is the dependency relation's distance[1] (Heringer *et al.*, 1980: 187; Hudson, 1995: 16).

Figure 1 furthermore illustrates that, in English, many heads and their dependents are immediately adjacent, - they have a dependency distance (DD) of 1. The Mean Dependency Distance (MDD) of a sentence/text is the sum of the individual dependency distances, divided by the number of dependencies that connect all the words of this sentence/text. Figure 2 illustrates the effect of extraction on MDDs.

1 Dependency distance can be quantified in different ways. Gibson (1998), for example, quantifies it in terms of new intervening discourse referents.

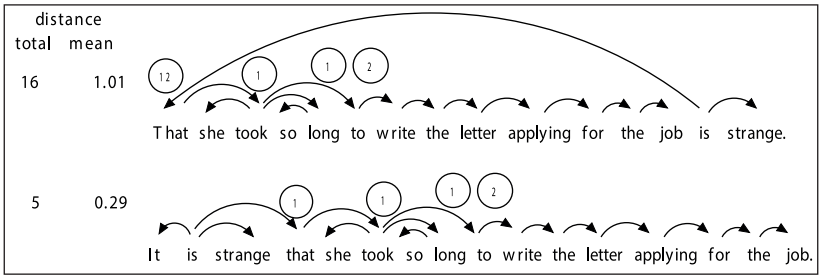

Figure 2. MDD of a sentence with default word order vs. extraction

Distance is an important property of dependencies because of its implications for language processing, i.e. how language is produced and understood. Constructing and interpreting sentences involves incrementally connecting words to arrive at meaning. This process consumes human or computational resources, i.e. it is "costly". Dependency distance has been shown to correlate with the cognitive cost of processing a linguistic dependency relation in terms of the memory cost required to keep track of incomplete syntactic (dependency) relations (Kimball, 1973; MacWhinney, 1987; Abney & Johnson, 1991; Gibson, 1991; Pickering & Barry, 1991; Lewis, 1993; Stabler, 1994; Hudson, 2010: 279); and the cost of connecting a new/incoming word to syntactically related ones. The computational cost of integrating a word into sentence structure has been shown to depend on the distance between a word and the most local head or dependent to which it attaches (Dependency Locality Theory DLT, Gibson, 2000). The DLT relates processing difficulty to dependency distance, i.e. it predicts that structures with longer dependencies are more difficult to process. The DLT can account for a number of processing complexity phenomena, e.g. the relative ease of subject- *vs.* object-extracted relative clauses; ambiguity resolution (in e.g. prepositional phrase attachment decisions), heaviness effects, and processing overload effects of multiple centre-embedded structures.

Considerations of parsing complexity have also been proposed to affect language production (Hawkins, 1994, 2004; Temperley, 2008). Synchronically and on the level of the individual speaker this seems to manifest itself mainly in phenomena of syntactic choice, e.g. default word order vs. extraction/extraposition (Temperley, 2008); diachronically Liu (2008) and Gildea and Temperley (2010) suggest dependency length minimisation may also play a role in the shaping of grammars, i.e. language evolution.

As dependency distance has implications for the cost of language processing, factors influencing dependency length need to be considered. Liu (2008) suggests that projectivity, or no crossing arches in the dependency graph of a sentence, influences dependency distance. Liu compared the mean distances of natural languages with those of artificial random languages, ones that allows crossing (non-projective) and ones that do not (projective). He found that non-projective

artificial grammars have the longest MDD, followed by projective artificial languages and natural ones. Liu interprets his results as showing "the usefulness of a no-crossing approach to dependency length reduction" (Liu, 2008:14; see also Gildea & Temperley, 2010: 307); in other words, he sees the reduced dependency distance of natural languages as a consequence of projectivity. Most well-formed strings in natural languages are projective (Marcus, 2007: 159).

If each word in a sentence has exactly one dependent, uniformly head-first or head-last structures yield shorter dependency distances than ones with pre- and post-dependents (Frazier, 1985; Hawkins, 1994; Rijkhoff, 1990). Predominantly head-first or head last-languages, such as Arabic and Japanese, should therefore have the shortest MDDs. Liu (2010) has shown that this is not the case because words can and do have more than one dependent (e.g. *consists* and *the* in the example in figure 1, *took* and *that* in the example in figure 2). If a word has more than one dependent, the grammar requires all dependents to point in the same direction, and there is a syntactic choice in terms of constituent order (e.g. a verb that has two prepositional dependents); placing the shorter dependent (phrase) closer to the head results in shorter dependencies. Hawkins (1994, 2004) reports that the preference of placing the shorter dependent closer to the head is found in head-first and head-last languages.

If a head has several dependents, placing all of them on the same side of the head creates a kind of "crowding" effect. German subordinate clauses, which are head final (rather than V2), illustrate that all dependents of the verbal head (*haben*) crowd to its left.

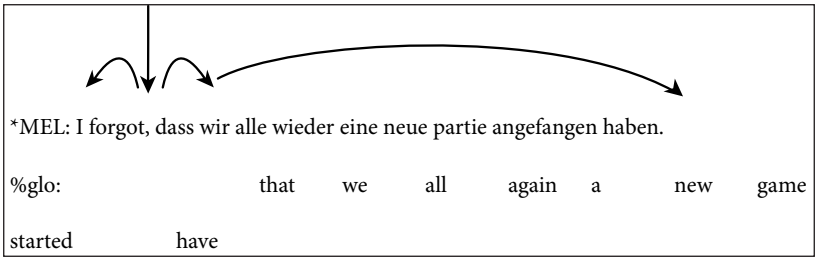

Figure 3. Crowding effect

In this case, there is no word order choice; if there was, placing some dependents to the left and some to the right of the verb would result in shorter dependencies. Generally, if a head has several dependents, balancing them on either side of the head results in shorter dependency distances (Temperley, 2008).

Languages that have a prevailing dependency direction but allow some short, dependent phrases to branch in the opposite direction have shorter MDDs than consistently same-branching languages like English (Dryer, 1992; Liu, 2010).

English is generally regarded as a predominantly head-first or right-branching language. In the English Penn Treebank, however, only 48.8% of the dependencies are head-first; German was found to be on average 54.5% head-first and Chinese 31.5% (Liu, 2010: 1571). Of the three languages we are looking at in this study, English has the best balance between left and right dependents and should therefore have the shortest MDD (followed by German and then Chinese). Section 2 below presents empirical support for this prediction.

So far we have established that dependency distance is a widely used measure for establishing the linear proximity of linguistically related words. DD can therefore be used to test whether there is indeed a preference for closely related words to be close together in monolingual language use (Gildea & Temperley, 2010). Other properties of dependency relations, the type of relationship they encode and their direction, were mentioned as possible factors influencing dependency distance. Most importantly, the effect that dependency distance seems to have on computational resources required for language processing and production was reviewed. Keeping track of long, incomplete dependencies is a burden on memory load, and the cost of linking a new word into sentence structure (by connecting it to its head and/or dependent) also seems to be influenced by the distance between the two (Gibson, 1998, 2000). In the next section we will look at dependency distance in the three languages involved in our data.

2. Dependency distance in English, German and Chinese

MDDs differ cross-linguistically. Although there is considerable variation in the type of language data analysed to date (spoken, written, formal, informal) and ways in which distance is measured and calculated,[2] there is a surprising amount of agreement as to which languages have short- and which ones have long-dependency distances.

Out of the three languages we are looking at, we anticipated English to have the shortest MDD, followed by German and Chinese. This is exactly what Liu (2008: 10) found: English (2.54) has a shorter mean dependency than German (3.35) and Chinese (3.66). Features of the three grammars that may account for this difference in dependency length will be looked at next.

The fact that English has fairly fixed word order and a prevailing dependency direction (head-first), but allows some short, dependent phrases to branch in the opposite direction seems to account for the short MDD of English (1.39, 1.49, 1.67 in Hiranuma's (1999), Eppler's (2010) and Wang and Liu's (2013) spoken data; 2.30 and 2.54 in the written data analysed by Gildea and Temperley (2010: 301) and Liu

2 Eppler (2005) and Hiranuma (1999) measure dependency distance in terms of the number of intervening words; Liu (2008, 2009, 2010) in terms of the difference between the word's position numbers. Liu (2009: 173) found the resulting difference in MDD to be to be small (1.81 *vs.* 1.89).

(2008: 12). In English, most words that are syntactically related are also adjacent; between 63% according to Collins (1996), 76% according to Pake (1998) and 78% according to Eppler (2010), but only slightly over 50% according to Liu (2008).

The mean distance between two syntactically related German words is longer than the mean distance between two related English words: 1.87 according to Eppler (2010), 3.07 according to Gildea and Temperley (2010), and 3.35 according to Liu (2008). The main syntactic reasons why German has a longer mean distance are; the generally freer word order in German, including scrambling, i.e. word order variation of argument noun phrases with respect to each other (examples 1a & b), with respect to adverbial phrases (example 2) or subjects (example 3)

(1a)	*Er hat ihr dieses Buch vielleicht gegeben.*
%glo:	he has her this book maybe given
(1b)	*Er hat dieses Buch vielleicht ihr gegeben.*
%glo:	he has this book maybe her given
(2)	*Er hat ihr vielleicht dieses Buch gegeben*
%glo:	he has her maybe this book given
(3)	*dass jeder den Geruch erkennt*
%glo:	that everybody this smell recognises

— the discontinuity between AUX/MOD and main verbs, i.e. the Verbalklammer

— the different word orders in German main (V2) and subordinate clauses (V final or SOV); According to Liu (2008: 17), German has more adjacent dependencies than both Chinese and English.

Chinese has the longest MDD, not only of the three languages we are looking at in this paper, but also of the 20 languages Liu (2008) compared: 2.85 in spoken news data (Wang & Liu, 2013:63), and 3.66 in written news data (Penn Chinese Treebank; Liu, 2008: 12). The fact that Chinese has fewer mixed (head-first/ head-last) dependencies than German and English and the fact that Chinese is an isolating language that marks e.g. tense, number and aspect with free (rather than inflectional) morphemes, has a significant influence on a) the number of word-word dependencies in a text and b) dependency length.

This brief cross-linguistic discussion of dependency length in English, German and Chinese has shown that rigidity of word order, consistency of dependency direction, and language type (isolating, inflecting) impact significantly on a language's MDD. Collins (1996), Pake (1998), Eppler (2010) and Liu (2008) have looked into the relationship between dependency length and adjacency. These preliminary findings are difficult to interpret and more work needs to be done on this in the future.

The comparison of MDDs in different data sets furthermore supports the idea that dependency distance is positively correlated with style (Liu Hudson & Feng, 2009: 171; Temperley, 2008). Casual speech has shorter distances than more formal speech and writing, even when of the same genre (i.e. news). The average difference in dependency length between spoken and written data in English, German and Chinese is approximately one word (1.02), with little variation between the three languages (Chinese 0.81, English 0.91 and German 1.34).

In the next sections we will look at bilingual data, data in which sentences are constructed from lexical items and grammatical structures from typologically different languages (English and German and English and Chinese). We will test whether syntactically related words from different languages also prefer to be close together, or whether long dependency distances facilitate code-switching (DH); i.e. we will investigate the effects of dependency distance on syntactic code-switching.

3. The data

The present paper is based on two bilingual corpora, a 9,023 word sample of a 93,235 word corpus of German/English (Eppler, 2003), and a 19,766 word corpus of Chinese/English speech (Wang & Liu, 2013). Both data sets were analysed in the same dependency theoretic framework (Hudson, 2007, 2010).

The German/English data was recorded in January and February 1993 among a close-knit network of members of the German-speaking Jewish refugee community who settled in London in the late 1930s. All speakers included in this sample are female and in their late sixties or early seventies. Their L1 is Austrian German; for all speakers the age of onset of their L2, British English, was during adolescence (15-21 years of age). In informal settings like the ones recorded, the participants use a bilingual mode of interaction sometimes called "Emigranto" (Eppler, 2010). Linguistically this mixed code is characterised by frequent switching at speaker turn boundaries and heavy intra-sentential code-switching. The audio data were transcribed in the CHAT/LIDES (LIPPS Group 2000) format and can be downloaded from <http://talkbank.org/data/LIDES/Eppler.zip>. The transcripts were manually annotated for word class, dependency type (see notation summary, appendix A), direction and distance. See table 1 below for a summary of the data.

The Chinese/English data (Wang & Liu, 2013) were audio-recorded from mainland China and Hong Kong TV or broadcasting programmes from June to September 2011. About 80% of the material is news, social news, and entertainment news; about 20% of the data are from interview programmes. Intra-sententially, code-switched sentences were selected from the data, transcribed and syntactically annotated to build a Treebank containing the following information: linear position of the head and the dependent in the sentence, word class, language and a selected number of dependency types (subject = SUBJ; object = OBJ; ATR = attribute; adverbial = AVDA). The MDD of the corpus and of individual dependency types were calculated from the Treebank using formulae proposed by Liu (2009). See table 1 for a summary of the data.

Table 1. Distribution of languages in the German/English and Chinese/English (Wang and Liu 2013) data

	German	English	Total	Chinese	English	Total
Word Tokens	5591	3432	9023	16267	3499	19766
Percentage	61.9	38.1	100	82.3	17.7	100

Table 1 shows that both data sets contain more word tokens from the speaker's L1 (German and Chinese) than from their L2 (English).

4. Dependency distance in "mixed" dependencies

"Mixed" dependencies are syntactic relations in which words A and B are from different languages. For mixed dependencies the main point of interest for this paper is whether dependency distance has an effect on code-switching. For mixed dependencies we might either expect:

— shorter distances that "counteract" the additional processing cost associated with code-switching for some speakers (but not others) (Dussias, 2001; Insurin Winford & De Boot, 2009). Code-switching is a mode of everyday interaction for the German/English bilinguals; we therefore do not assume code-switching to consume additional processing resources for these speakers.

— a dependency distance between the mean distances for English and German and English and Chinese monolingual dependencies, because syntactic dependency properties of both languages are involved.

— longer mixed dependency distances, if the influence of a word's language on that of its dependent decreases with increased distance i.e. if priming[3] effects decline over distance. In activation-based frameworks, the activation level of a word (and its language-specific morpho-syntactic properties) will decay with distance. Structural integration involves reactivating the word to a target threshold level so that aspects that are relevant to its integration (e.g. agreement features, positional information) can be retrieved from memory. This reactivation is not only costly, but may also be incomplete; information about a word's number, gender, case or language may degrade and/or be lost. In this case, the longer the distance, the more likely we are to encounter an "other" language dependent, i.e. a code-switch.

The distance of mixed dependencies with a German head is expected to be longer than the distance of mixed dependencies with an English head because monolingual German dependencies are longer than English ones and heads may influence dependency distance more than dependents. Placing dependents on either side of the

3 A given word's language activates mental pathways that enhance access to subsequent words of the same language in memory.

head only results in shorter dependency distances if the dependent phrases are short (see section 1). Verbs are main heads and German verbs are involved in construction types that involve a combination of long left- and right-dependents (see section 1).

We expect the distance of mixed dependencies with a Chinese head to be longer than the distance of mixed dependencies with an English head, because the MDD of Chinese is longer than that of English.

The empirical findings presented in the next sections will reveal which of these possibilities are supported by the German/English and Chinese/English data.

5. Results on MDD in German/English and Chinese/English

Table 2 presents the MDDs for monolingual German, monolingual English and mixed dependencies with German and English heads, respectively.

Table 2. Mean distances of monolingual and mixed German/English dependencies

	German	English	Average
Monolingual	1.87	1.49	1.68
Mixed with head	1.85	2.26	2.06

These findings support the proposition that monolingual German dependencies will be longer than monolingual English ones (made on the basis of the word order properties of the two languages in section 2); and findings by Liu (2008) and Gildea & Temperley (2010).

The mean distances of mixed dependencies with a German head do not support the ideas that heads have a significant effect on dependency distance, and that mixed dependencies counter-balance a potentially greater processing load of bilingual utterances with short distances. This finding may, however, indicate that with German verbs, the word class that is assumed to trigger the most substantial increase in dependency distance through bi-directional long-distance dependencies, are infrequently involved in mixed dependencies.

The mean distance of mixed dependencies with an English head suggests that English words enter into "looser", literally more remote, syntactic relations with German dependents. We therefore expect English words to "head" more dependency relations that are characterised by long distances, e.g. adjunct, extractee and extraposee relations; and we expect German dependents of English heads to be more frequently located at the clause periphery. This idea would tie in well with the literature on code-switching. Treffers-Daller (1994) first noted a high propensity of switching for "dislocated constituents" in her French/Dutch data. Muysken (2000) subsequently adopted the idea that code-switching is favoured in clause peripheral positions as one of four primitives of code-switching.

Interpreted within an activation-based framework, the findings on mixed dependencies seem to indicate that English words create higher activation levels for the "language" property than German words. This could mean that German (the L1) is functioning as a kind of base language in the speech of the German/English bilinguals, and English (the L2) is more marked.

The highly significant difference between monolingual and mixed dependency distances ($X2 = 18.6$, $df = 1$, $p < 0.001$) seems to support the idea that dependency distance affects code-switching. Long dependency distances seem to increase the likelihood of another language dependent, i.e. a code-switch. The Distance Hypothesis (DH) proposes that greater dependency distance increases the chances of code-switching.

The DH suggests that activation levels of related words may be a motivating factor behind code-switching. Both the head and the dependent of a dependency relation need to be – or be made – active at the point in language production/processing when the dependency between them is being established. Activation levels of words (and their properties), however, decay as intervening words are being processed/produced and integrated into sentence structure. The processing load of long distance dependency relations is therefore high (Gibson, 1998, 2000) and the priming effect the head and the dependent have on each other low. Mixed dependencies/code-switches may be the result of distance because the influence of a word's language on that of its dependent decreases with increased distance.

The Distance Hypothesis is a syntactic processing hypothesis, i.e. it combines features of grammar (dependency relations) and psycholinguistic processes of speech production (dependency distance). Evidence in its support would therefore potentially shed light on both grammatical and psycholinguistic aspects of code-switching.

The recent analysis of a Chinese/English code-switched corpus in terms of dependency distance and direction revealed interesting similarities and differences between the German/English (table 2) and Chinese/English data (Wang & Liu, 2013; table 3).

Table 3. Mean distances of monolingual and mixed Chinese/English dependencies

	Chinese	English	Average
Monolingual	2.85	1.67	2.26
Mixed with head	3.54	2.81	3.18

The Chinese/English data show that monolingual Chinese dependencies are longer than English ones ($p < 0.005$). This was expected from the morphological and word-order properties of the two languages (section 2) and supports Liu (2008) and Liu, Hudson and Feng (2009).

The average MDD of mixed dependencies is longer than that of monolingual dependencies, and the average MDDs of mixed dependencies is longer than the

MDDs of both English and Chinese monolingual dependencies ($p < 0.05$). These core findings are almost identical to those from the German/English data (the exception are the extremely long, mixed dependencies with an English head in the German/English data).

The mean distance of mixed dependencies with an English head and a Chinese dependent is also longer than that of monolingual English dependencies in the Chinese/English data, but the difference is not quite as marked ($p < 0.05$ *vs.* 0.001).

The mean distance of mixed dependencies with a Chinese head and an English dependent is longer than that of monolingual Chinese dependencies ($p < 0.001$). This is different to what we found in the German/English data where mixed dependencies with an L1 head are marginally shorter than monolingual L1 dependencies.

The mean distance of mixed dependencies with a Chinese head and an English dependent is much longer than that of mixed dependencies with an English head and a Chinese dependent ($p < 0.05$). However, the increase in MDD is larger from monolingual English to Chinese mixed with an English head (+1.14) than that from monolingual Chinese to English mixed with a Chinese head (+ 0.69). The pattern is parallel to what we found in the German/English data, where the mean distance between monolingual English and mixed dependencies with an English head increases by (+ 0.77). If heads affect dependency distances, it is interesting to note that heads from the speakers' L1s (German and Chinese) hold their dependents "tighter" than L2 heads; the respective L1s may be functional as a "base" language for both sets of bilinguals.

To summarise, the comparison of tables 2 and 3 have shown that

1. MDDs are cross-linguistically different, with English having the shortest MDD, followed by German and Chinese
2. monolingual dependencies in mixed corpora are slightly (but not significantly) different to those found in comparable monolingual corpora
3. the average MDDs of mixed dependencies are longer than those of monolingual dependencies

The analysis of the Chinese/English data conducted by Wang and Liu (2013) correspond to the general patterns found in the German/English data. Most importantly, greater dependency distance also seems to increase chances of code-switching in Chinese/English bilingual speech. The findings from a typologically very different language pair and data set therefore substantiate the idea that code-switching is more likely in syntactic relations with long dependency distances, i.e. the Distance Hypothesis (DH).

In the next sections we will compare individual dependency types from the two data sets in terms of distance.

6. Dependency distance of individual dependency types

The following three sub-sections focus on the dependency distances of selected syntactic relations in the German/English data (subjects, objects, other complements, adjuncts, x-comps/sharers and extractees/extraposees). Cross-references to the Chinese/English data will be made where possible. The Chinese/English data were analysed for four syntactic relations (subjects, objects, attributes and adverbials – both of which are analysed as adjuncts in the German/English data). Section 6.1 compares monolingual dependencies with each other, section 6.2 compares monolingual L1 dependencies with mixed dependencies with an L1 head, and section 6.3 compares monolingual L2 dependencies with mixed dependencies with an L2 head. The findings support the main idea outlined in the previous section, the Distance Hypothesis, and some of the constraints on code-switching proposed in the literature.

6.1 Monolingual dependencies

Table 4 illustrates how individual dependency types contribute to the average dependency distance of 1.87 for monolingual German dependencies and 1.49 for monolingual English dependencies in the German/English data.

Table 4. MDDs (and frequencies) of monolingual German and monolingual English dependency types; s- subject; o- object; r - x-comp/sharer[4]; c-complement (of words other than verbs); a - adjunct; x - extractees/extraposee; < & > indicates left- and right dependent

	s <	> s	> o	o <	> r	r <	>c	> a	a <	x <	total
G-G	1.54 (142)	1.07 (45)	1.78 (54)	1.83 (36)	2.64 (69)	1.07 (69)	1.65 (155)	2.1 (100)	1.37 (86)	3.16 (19)	1.68 (754)
E-E	1.07 (130)	- (7)	1.5 (82)	- (0)	1.53 (93)	- (0)	1.22 (130)	2.26 (72)	1.38 (44)	- (3)	1.06 (596)

The column entries of table 4 demonstrate that different dependency types have different mean distances (Liu, Hudson & Feng, 2009: 170); the rows show that MDDs differ cross-linguistically (Liu, 2008; Liu Hudson & Feng, 2009) and that the German/English bilinguals' word orders in terms of Stellungsfelder seems to be intact (there are no English left-dependent objects and x-comps). The mean dependency distances that differ most significantly between German and English are subjects and x-comps/sharers. These differences are caused by the Verbalklammer[5] and the subjects of clause-final finite verbs, which are at almost opposite ends of subordinate clauses. Gildea & Temperley (2010: 301) also found

4 Sharer is the Word Grammar (Hudson, 2010) term for a verb complement that shares its subject with the head of the sentence.

5 The Verbalklammer is particularly frequent in corpora of spoken German where reference to past time is made with the present perfect (*Ich habe ... geshen*) rather than the simple past (*Ich sah*).

that verb position contributes to the longer dependency distances in German, but stress that it is not specifically the distance from subject to verb that results in this effect. Given that subjects tend to be short and can frequently be placed on either side of the verb in German, this finding is in line with the interrelation between dependency direction and distance, as discussed in section 1.

From table 5, the biggest difference in mean distances between monolingual Chinese and English lies in the subject relation ($p<0.001$).

Table 5. The MDD and frequencies of 4 major Chinese and English monolingual grammatical relations

	Subj	Obj	Atr	Adva	Average
C-C	2.55 (940)	2.74 (849)	1.59 (1505)	2.45 (3039)	2.33
E-E	1.41 (130)	1.65 (91)	1.17 (296)	1.92 (104)	1.54

Chinese prepositional constructions, such as *bei, ba, jiang* or *ge* and the complement of *di,* which are used as adverbials, must follow the subject but precede the modified verb; this increases the dependency distance between the subject and the head/root of Chinese sentences, as in example (4).

(4) wo ba ta dang pengyou.
I PREP him treat friend
"I treat him as my friend."

Figure 4 shows that the dependency distance of the subject relation between *wo* and the verb *dang* is 3 in Chinese; the dependency distance of the English translation *I treat* is only 1.

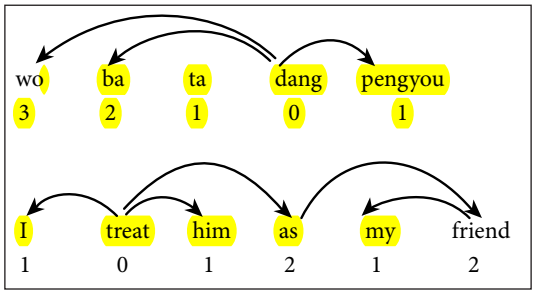

Figure 4. The dependency structure of example 4 and its English translation

Wang and Liu (2013) also found longer MDD of Chinese objects in comparison with English ones ($p<0.001$). Tense is realised by inflectional morphology in English; in Chinese, tense is usually handled by function words that separate the object and the head. In example (5), illustrated in fig. 5, the dependency distance between the

object *book* and its head *bought* is 2. In the Chinese translation *wo mai le zhe-ben shu* the dependency distance between *mai* and *shu* is 4, because the complement of the classifiers *zhe* and *ben* and the perfect-tense auxiliary *le* intervene between the object and it's verbal head.

(5) *wo mai le zhe-ben shu.*
I buy AUX this-CL book
"I bought the book."

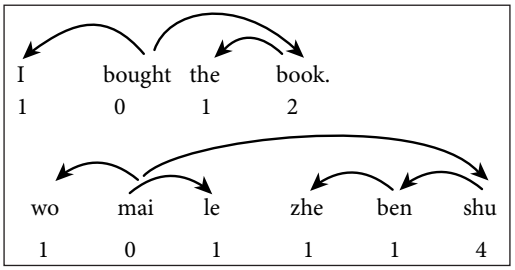

Figure 5. The dependency structures of example 5 and its translation

There are similar cases in German where a verb particle can separate the verb from their object, as in *Er lief den Hügel hinauf* "He ran up the hill". Constructions like these raise the question what size linguistic unit we should analyse.

6.2 Monolingual L1 and mixed dependencies with an L1 head

Table 2 showed that the mean distance of mixed dependency relations with a German head is marginally shorter than the mean distance of monolingual German dependencies (0.85 to 0.87). Table 6, however, reveals that the distances for most mixed grammatical functions (subjects, adjuncts, pre-dependent sharers and post-dependent objects) are longer than their monolingual German equivalents.

Table 6. MDDs and frequencies of monolingual German and mixed syntactic dependencies with a German head

	s<	>s	>o	o<	>r	r<	>c	>a	a<	>x	x<	total
G-G	1.54 (142)	1.07 (45)	1.78 (54)	1.83 (36)	2.64 (69)	1.07 (15)	1.65 (155)	2.1 (100)	1.37 (86)	- (3)	3.16 (19)	1.68 (754)
G-E	1.7 (10)	1.5 (2)	2.38 (29)	1.5 (20)	1.95 (61)	1.29 (7)	1.1 (309)	3.9 (38)	1.52 (27)	1.33 (3)	3.07 (14)	2.06 (525)

The slightly shorter mean distance of mixed dependencies with a German head (in comparison with monolingual German dependencies) is only attributable to three dependency types: complements, post-dependent sharers and pre-dependent objects. Out of these three, it is the very large number of borrowed English noun complements of German determiners that brings the mean distance down.

This result also tells us something about the syntactic structure of mixed complement relations with an English dependent noun: they are hardly ever pre-modified. A lot of the English predicative adjectives are very close to their German head; and so are the English objects that depend on German clause final/SOV verbs. The fact that English post-dependent adjuncts are almost three times as far away from their German head as monolingual post-dependent adjuncts seems to support Treffers-Daller (1994), Mahootian and Santorini (1996) and Muysken (2000), i.e. that code-mixing is favoured in adjoined peripheral positions, as in example (6).

(6) *MEL: *nein # ich bin draussen # **as per usual**.*
%tra: no # I am out
Jen2.cha: line 185.

In table 7, the MDD of adverbials (5.65) in C-E dependencies is much longer than that in Chinese dependencies ($p<0.001$).

Table 7. MDDs and frequencies of 4 major Chinese monolingual grammatical relations compared to mixed ones with a Chinese head and an English dependent

	Subj	Obj	Atr	Avda	Average
C-C	2.55 (940)	2.74 (849)	1.59 (1505)	2.45 (3039)	2.33
C-E	2.7 (161)	2.85 (310)	1.48 (43)	5.65 (54)	3.17

In example (7) displayed in figure 6, the switched English adverbial *suddenly* has a distance of 2 because it is located at the clause periphery. In Chinese, adverbials always precede their verbal head, but they can precede or follow the subject. In the corresponding monolingual Chinese sentence the adverbial *turan* follows the subject and the dependency distance of *turan-mingbai* is only 1.

(7) *Suddenly, wo mingbai le ta de yisi.*
I understand AUX he AUX meaning
"Suddenly, I understand what he said."

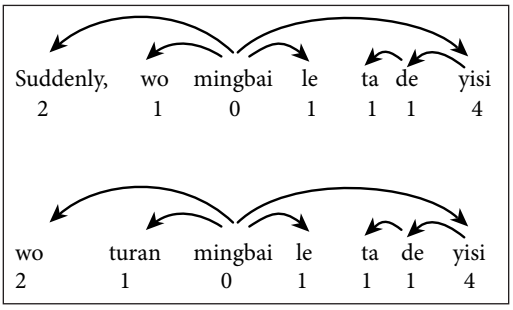

Figure 6. The dependency structures of example 7 and its translation.

The Chinese/English data furthermore contain an interesting example that involves the extraposition of an English subject. In example (8) displayed in figure 7, the English subject *fans* is left-dislocated from its default position preceding *zhuanmen*. Because *fans* is dislocated to the left clause periphery, and because the prepositional phrase must precede the verb in Chinese, the distance between *fans* and its Chinese head *jisong* is 6; in the corresponding monolingual Chinese sentence, the distance between the Chinese subject *fensimen* and its head *jisong*, on the other hand, is only 2.

(8) *Fans weile xiang ta zhijing zhuanmen jisong xianhua.*
Fans in order to to him pay their respects specially posted flowers
'In order to pay their respects to him, fans specially posted flowers.'

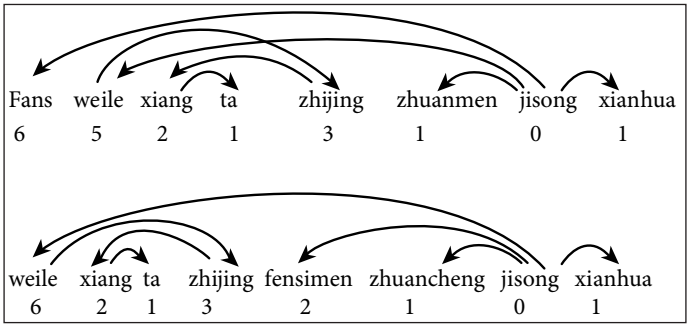

Figure 7. The dependency structures of example 8 and its Chinese corresponding sentence

In section 4 we suggested that the mean distance of mixed dependencies with a German head might be marginally shorter than the mean distance of monolingual German dependencies because the word class that is assumed to increase dependency distance through a change in dependency direction, i.e. German verbal heads, is infrequently involved in mixed dependencies. An analysis of all German verbs in the German/English corpus revealed that members of this word class do function as heads in mixed dependencies. A separate test performed on verb types (main *vs.* AUX/MOD) showed that overall German verbs are not significantly less frequently involved in mixed dependencies than monolingual ones ($p=0.112$). The same holds true for German main verbs ($p=0.192$). German auxiliaries and modals, however, are significantly more frequently involved in mixed dependencies than English ones ($p < 0.001$). This finding is interesting as AUX / MOD are frequently in V2 position in German, which often coincides with the placement of verbs in SVO structures such as English. German AUX and MOD are therefore placed in congruence sites (Sebba, 1998). Congruence sites / equivalent surface word orders have been identified as factors that facilitate code-switching (Muysken, 2000).

6.3 Monolingual L2 and mixed dependencies with an L2 head

To explain the results presented in table 2 we hypothesised that the influence of a word's language on that of its dependent may decrease with increased distance (because of an increased memory and integrations cost (Gibson, 1998, 2000) and a degraded priming effect). Mixed dependencies would therefore be the result of distance. As a consequence of their long dependency distances, code-switches were also expected to be more frequently located at the clause periphery.

More specifically, on the basis of the mean dependency distances in the German/English data (table 2) we proposed that English heads may enter into "looser", literally more remote, syntactic relations with German dependents. We predicted that English words to "head" more dependency relations that are characterised by long distances, e.g. adjunct, extractee and extraposee relations, and expected German dependents of English heads to be more frequently located at the clause periphery (*cf.* Treffers-Daller, 1994). This is what we find in the data.

Table 8. MDDs and frequencies of selected monolingual English and mixed syntactic functions with an English head

	>c	s <	> a	a <	> r	> o	> x	x <	Total
E-E	1.22 (130)	1.07 (137)	2.26 (116)	1.38 (116)	1.53 (93)	1.5 (82)	- (1)	- (3)	1.64 (596)
E-G	1.84 (45)	1.9 (11)	2.33 (55)	1.78 (55)	3.12 (7)	1.18 (18)	1.45 (7)	4.5 (15)	2.06 (165)

Focusing on the mean distances of individual dependency types in table 8 we notice that ALL mixed dependencies with an English head (apart from objects) are longer than their monolingual English counterparts (this is unlike the mean distances of monolingual German and mixed grammatical relations with a German head (table 6). Table 8 further illustrates that all dependency relations that yield a significantly higher number of mixed tokens than monolingual ones (German adjuncts, extractees), are further away from their English heads than their English counterparts. This finding supports the Distance Hypothesis, which proposes that code-switching is more likely in long dependency relations.

Table 8, for example, shows that the adjunct relation is very popular for switching between an English head and a German dependent.

(9) *LIL: *die xx hat es # **in high heaven** gelobt.*
%glo: xx has it # praised
Jen2.cha, line 1570

(10) *MEL: *als kind **I didn't like anything** aber **I love food** .*
%tra: as a child I didn't like anything but I love food
Jen2.cha, line 2058

Note that the pre-adjunct in (10) is also extracted; it is moved out of its default word order position and moved to the left clause periphery, which increases its dependency distance by 4.

Example (11) illustrates a German long-distance (distance = 8) extraction.

(11) *MEL: *was die Dorit wieder geschmissen hat, **I** [/] **I would have liked**.*
%glo: what the Dorit again thrown has,

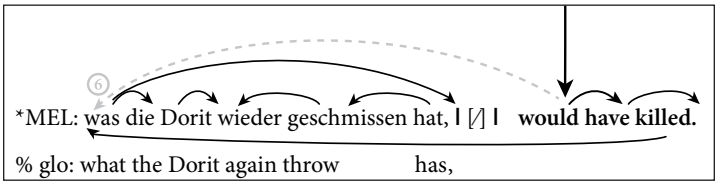

Figure 8. The dependency structures of example 11

It appears that for stylistic reasons (Temperley, 2008), speaker MEL increases the distance of a mixed dependency relation from zero to eight in the above example.

The hypothesis that L2 heads predominantly enter less intimate and longer syntactic relations with L1 dependents is also supported by the Chinese/English data (table 9).

Table 9. MDDs of 4 major grammatical relations in monolingual English and mixed dependencies with an English head and a Chinese dependent

	Subj	Obj	Atr	Avda	Average
E-E	1.41 (130)	1.65 (91)	1.17 (296)	1.92 (104)	1.54
E-C	2.75 (87)	2.88 (32)	1.67 (446)	2.07 (311)	2.55

In this direction, i.e. hL2-dL1, the findings for switched adjuncts from the Chinese/English data are very similar to the finding from the German/English corpus. There are significantly (<0.001) more switched Chinese adjuncts than subjects and objects. This is in stark contrast with the significantly (<0.001) larger number of switched Chinese subjects and objects in the other direction hL1-dL2. The mean dependency distance of the switched Chinese adjuncts is longer than that of monolingual English ones. The increase in mean dependency distance between monolingual and mixed dependencies, however, is bigger in all other dependency types analysed in the Chinese/English corpus, i.e. attributes, objects, and subjects. The distance between the Chinese subject *tamen* and its head is 2 in example (12), illustrated in fig. 9, because the adverbial *tiantian* precedes the verb. In its English translation the DD between *they* and *send* is only 1.

(12) *Tamen tiantian send E-mails.*
They everyday send
"They send E-mails everyday."

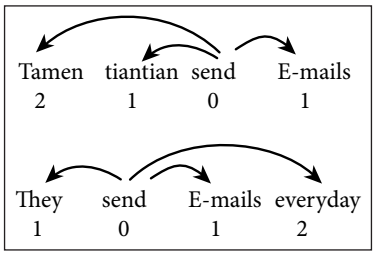

Figure 9. The dependency structures of example 13 and its translation

In example (13) displayed in figure 10, the distance of *understand-yiqie* is 5; in its English translation the distance of *understand-everything* is only 1.

(13) *I fully understand ni gaosu ta de yiqie*
you tell him AUX everything
'I fully understand everything that you tell him.'

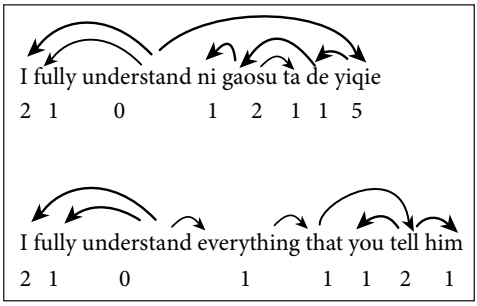

Figure 10. The dependency structures of example 6 and its translation

The hypothesis that greater dependency distance of syntactic relations increases the chances of code-mixing appears to apply particularly to mixed syntactic relations with an L2 head. Mixed grammatical functions with an L2 head seem to pose a particular processing complexity for the German/English and the Chinese/English bilinguals alike, and the activation of L2 heads appears to decay more rapidly in long-distance dependencies than that of L1 heads. This seems to render the retrieval of features of the L2 head (e.g. its language) more difficult from memory and lead to the significantly larger number of mixed, long-distance syntactic relation with an L2 head in both corpora. The results presented in table 8 and the findings from

the Chinese/English Treebank (table 9) also support the notion that code-mixing is favoured in peripheral and adjoined positions (Treffers-Daller, 1998; Muysken, 2000).

7. Summary and conclusion

We started from a well-established principle of monolingual language comprehension and production, that closely related words tend to be close together in the sentence, and suggested that this may be different in bilingual language use. The analysis of a 9,023 word corpus of German/English mono- and bilingual speech indeed revealed that mixed dependencies have a longer MDD than monolingual ones. This led to the formulation of the Distance Hypothesis and a set of specific hypotheses on the syntactic behaviour of linguistic elements in the German/English data.

The central syntactic processing claim to emerge from the quantitative analysis is that the influence of a word's language on that of its dependent seems to decay with the number of words intervening between it and its dependent. The rationale behind this idea is that both the head and the dependent in a syntactic relation need to be, or be made, active at the point in the production process when the dependency between them is being established; if the head and the dependent are far apart, the priming effect of the language of the first word will have decayed with time/distance; therefore the longer the dependency link, the less the priming influence, and the more likely a change in language. This argument is supported by independent findings that long distance dependency relations are difficult to process, both in terms of memory load and integration cost (Gibson, 2000).

In a 19,766 word corpus of a typologically very different language pair, Chinese/English, mixed dependencies also have longer mean dependency distances than monolingual ones. The analysis of individual grammatical functions in both corpora revealed that, with one exception in the Chinese/English corpus and three in the German/English data, all mixed dependency relations are, on average, longer than the corresponding monolingual ones. Both corpora contain considerable numbers of long-distance mixed adjuncts, and in the German/English data L2 heads tend to enter into "looser", literally more remote, syntactic relations with L1 dependents, i.e. syntactic relations that are not essential for building sentence structures (adjunction, extraction and extraposition). When L1 subjects and objects of L2 verbs have long dependency distances, they are also frequently switched in the Chinese/English data. In languages where the root verb tends to occupy a central sentence position, such as English, German and Chinese, long distance dependents will be located at the clause periphery. That code-switching is favoured in clause-peripheral positions has been established in bilingualism research (Treffers-Daller, 1994; Muysken, 2000). The DH however, captures this notion on a more general syntactic processing level.

The results from the German/English and Chinese English data are promising. But in order to establish dependency distance between syntactically related units as a

viable motivator for code-switching and – more generally – for contact-induced language change, the distance hypothesis will have to be tested on other bilingual corpora and with controlled psycholinguistic experiments.

References

Abney, S. P.; Johnson, M. "Memory requirements and local ambiguities of parsing strategies". *Journal of Psycholinguistic Research*, 20, 1991, 233-250.

Bentahila, A.; Davies, E. E. "The syntax of Arabic - French code-switching". *Lingua*, 59, 1983, 301-30.

Collins, M. J. "A new statistical parser based on bigram lexical dependencies". *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics. University of California, Santa Cruz, 24-27 June 1996,* 184-191.

Dryer, M.S. "The Greenbergian word order correlations". *Language*, 68, 1992, 81-138.

Eppler, E. German/English data base [Online]. 2003. <http://talkbank.org/data/LIDES/Eppler.zip> (Access date: 15 February, 2013).

Eppler, E. "Word Grammar and syntactic code-mixing research". In: Sugayama, K.; Hudson, R.A. (ed.). *Studies in Word Grammar: New Perspectives on a Theory of Language Structure*. London, New York: Continuum, 2006, 117-139.

Eppler, E. *Emigranto. The syntax of German/English code-switching*. Vienna: Wilhelm Braumüller, 2010.

Eppler, E.; Davies, J. "Linguistic transfer - bridges between bilinguals' languages". In: Coelsch-Foisner, S.; Markus, M.; Schendl, H. (ed.). *Transfer in English Studies*. Vienna: Braumueller, 2012, 239-258.

Frazier, L. "Syntactic complexity". In: Dowty, D.; Karttunen, L.; Zwicky, A. (ed.). *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. Cambridge, UK: Cambridge University Press, 1985, 129-189.

Gibson, E. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. [PhD Dissertation]. Pittsburgh PA, Carnegie Mellon University, 1991.

Gibson, E. "Linguistic complexity: Locality of syntactic dependencies". *Cognition*, 68, 1998, 1-76.

Gibson, E. "The dependency locality theory: A distance-based theory of linguistic complexity". In: Marantz, A.; Miyashita, Y.; O'Neil, W. (ed.). *Image, Language, Brain*. Cambridge, MA: MIT Press, 2000, 95-126.

Gildea, D. & Temerpley, D. "Do grammars minimize dependency length?". *Cognitive Science*, 34, 2010, 286-310.

Gumperz, J. J.; Hernandez-Chavez, E. "Cognitive aspects of bilingual communication: Language use and social change". In: Whiteley, W. H. (ed.). Oxford: Oxford University Press, 1971, 111-25.

Hawkins, J. A. *A Comparative Typology of English and German: Unifying the Contrasts*. London and Sydney: Croom Helm, 1986.

Hawkins, J. A. *A Performance Theory of Order and Consistency*. Cambridge: Cambridge University Press, 1994.

Hawkins, J. A. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press, 2004.

Heringer, H. J.; Strecker, B.; Wimmer, R. *Syntax: Fragen-Lösungen-Alternativen*. München: Wilhelm Fink Verlag, 1980.

Hiranuma, S. "Syntactic difficulty in English and Japanese: A textual study". *UCL. Working Papers in Linguistics*, 11, 1999, 309-322.

Hudson, R. "Measuring syntactic difficulty". 1995. [Online]. <http://www.phon.ulc.ac.uk/home/dick/difficulty.htm>. (Access date: 6 July, 2008).

Hudson, R. *Language Networks: The New Word Grammar*. Oxford: Oxford University Press, 2007.

Hudson, R. *An Introduction to Word Grammar*. Cambridge: Cambridge University Press, 2010.

Kimball, J. "Seven principles of surface structure parsing in natural language". *Cognition*, 2(1), 1973, 15-47.

Lewis, R. *An Architecturally-based Theory of Human Sentence Processing*. [PhD Dissertation]. Pittsburgh, PA, Carnegie Mellon University, 1993.

Liu, H. "Dependency distance as a metric of language comprehension difficulty". *Journal of Cognitive Science*, 9(2), 2008, 1-24.

Liu, H.; Hudson, R. A.; Feng, Z. "Using a Chinese Treebank to measure dependency distance". *Corpus Linguistics and Linguistic Theory*, 5(2), 2009, 161-174.

Liu, H. "Dependency direction as a means of word-order typology: A method based on dependency treebanks". *Lingua*, 120, 2010, 1,567-1,578.

MacWhinney, B. "The competition model". In: MacWhinney, B. (ed.). *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 1987, 249-308.

Marcus, S. "Hidden Grammars". In: Marcus, S. (ed.). *Words and Languages Everywhere*. Milano: International Scientific Publisher, 2007, 155-162.

Mahootian, S.; Santorini, B. "Code-Switching and the Complement/Adjunct Distinction". *Linguistic Inquiry*, 27, 1996, 464-79.

Muysken, P. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press, 2000.

Pake, J. *The Marker Hypothesis: A constructivist theory of language acquisition*. [PhD dissertation]. Edinburgh University, 1998.

Pickering, M.; Barry, G. "Sentence processing without empty categories". *Language and Cognitive Processes*, 6, 1991, 229-259.

Pfaff, C. "Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English". *Language*, 55, 1975, 291-318.

Poplack, S. "Sometime I'll start a sentence in Spanish y termino en Espanol: Toward a typology of code-switching". *Linguistics*, 18, 1980, 581-618.

Rijkhoff, J. "Explaining word order in the noun phrase". *Linguistics*, 28, 1990, 5-42.

Sebba, M. "A congruence approach to the syntax of code-switching". *International Journal of Bilingualism*, 2(1), 1998, 1-19.

Stabler, E. P. "The finite connectivity of linguistic structures". In: Clifton, C.; Jr., Frazier, L.; Rayner, K. (ed.). *Perspectives on Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum, 1994, 303-336.

Temerpley, D. "Dependency length minimization in natural and artificial languages". *Journal of Quarterly Linguistics*, 15, 2008, 256-282.

Treffers-Daller, J. *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: De Gruyter, 1994.

Wang, L.; Liu, H. "Syntactic variations in Chinese-English code-switching". *Lingua*, 123, 2013, 58-73.

EVA DURAN EPPLER

Is Reader in English Language and Linguistics at the University of Roehampton, London, UK. In search of a transcription system for her corpus of German/English bilingual speech she got involved with the *Language Interaction in Plurilingual and Plurilectal Speakers* (LIPPS) group, where she met Maite. Together they published the *Language Interaction Data Exchange Coding Manual* (The LIPPS Group 2000). Since then Eva Duran Eppler has mainly been working on structural and processing aspects of bilingual speech. Recently she has published *Emigranto: the syntax of German English Code-switching*, based on the LIDES data; and authored and edited *English in its Social Context and English Words and Sentences* (CUP).