Research Note

The International Journal of Bilingualism

International Journal of Bilingualism
1–20

DOI: 10.1177/1367006917729543
journals.sagepub.com/home/ijb


Code-switching within the noun phrase: Evidence from three corpora

M. Carmen Parafita Couto
Leiden University, the Netherlands

Marianne Gullberg
Lund University, Sweden

Abstract

Aims and objectives/purpose/research questions: This study aims to improve our understanding of common switching patterns by examining determiner–noun–adjective complexes in code-switching (CS) in three language pairs (Welsh–English, Spanish–English and Papiamento–Dutch). The languages differ in gender and noun–adjective word order in the noun phrase (NP): (a) Spanish, Welsh, and Dutch have gender; English and Papiamento do not; (b) Spanish, Welsh, and Papiamento prefer post-nominal adjectives; Dutch and English, prenominal ones. We test predictions on determiner language and adjective order derived from generativist accounts and the Matrix Language Frame (MLF) approach.

Design/methodology/approach: We draw on three publicly available spoken corpora. For the purposes of these analyses, we re-coded all three datasets identically. From the three re-coded corpora we extracted all monolingual and mixed simplex NPs (DetN) and complex NPs with determiners (determiner–adjective–noun (DetAN/NA)). We then examined the surrounding clause for each to determine the matrix language based on the finite verb.

Data and analysis: We analysed the data using a linear regression model in R statistical software to examine the distribution of languages across word class and word order in the corpora.

Findings/conclusions: Overall, the generativist predictions are borne out regarding adjective positions but not determiners and the MLF accounts for more of the data. We explore extra-linguistic explanations for the patterns observed.

Originality: The current study has provided new empirical data on nominal CS from language pairs not previously considered.

Significance/implications: This study has revealed robust patterns across three corpora and taken a step towards disentangling two theoretical accounts. Overall, the findings highlight the importance of comparing multiple language pairs using similar coding.

Keywords

Bilingualism, code-switching, corpora, noun phrases, Dutch, English, Papiamento, Spanish, Welsh

Corresponding author:
Maria Carmen Parafita Couto, Leiden University Center for Linguistics, Witte Singel-complex, van Wijkplaats 3, 2311 BX Leiden, The Netherlands.
Email: m.parafita.couto@hum.leidenuniv.nl

Introduction

Studies of code-switching (CS) – or the use of multiple languages in one utterance – cover a range of languages and linguistic domains (e.g. Bullock & Toribio, 2009; Deuchar, 2012). However, many issues remain poorly understood. For instance, the question of what does and does not occur in CS remains difficult to assess given the poor accessibility of comparable data for meta-analysis. Theoretical accounts make different predictions about the location and the directionality of switches but the empirical evidence remains contradictory.

The current study aims to improve our understanding of common switching patterns by examining noun phrases (NPs; determiner–noun–adjective complexes) as a conflict domain in three language pairs. We examine three conversation corpora whose languages have different properties in the nominal domain, namely Welsh–English, Spanish–English, and Papiamento–Dutch. Importantly, the languages differ with regard to gender and noun–adjective word order. Spanish, Welsh and Dutch have gender whereas English and Papiamento (a creole spoken in the Dutch Antilles (Gordon, 2005; Kouwenberg & Murray, 1994)) do not. Spanish, Welsh and Papiamento prefer post-nominal adjectives, whereas Dutch and English prefer prenominal adjectives. These language pairs thus allow us to test predictions about NP-switching in cases where word order differs across languages and where there are asymmetries regarding gender marking.

To guide the analyses we consider predictions derived from two theoretical traditions, namely generativist accounts (Cantone & MacSwan, 2009; Liceras, Fuertes, Perales, Pérez-Tattam & Spradlin, 2008) and the Matrix Language Frame (MLF) approach (Myers-Scotton, 1993). We explore how they can account for the relationship between determiners and nouns, and the complex relationships in adjectivally modified NPs. To do so, we compare these constructions in three identically coded corpora.

Background

In early analyses of CS the notion of grammatical surface equivalence played an important role. Following Pfaff (1979), Poplack (1980; 1981) postulated two famous constraints. The *free morpheme constraint* states that switching is not possible between a bound morpheme and its host. The *equivalence constraint* further restricts CS to locations in the clause where surface structures of the languages match, prohibiting switching where the surface orders differ. Many studies ensued, often reporting counterexamples to the constraints (Azuma, 1993; Bentahila & Davies, 1983; Berk-Seligson, 1986; Cantone & Müller, 2008; Di Sciullo, Muysken & Singh 1986; Jake, Myers-Scotton & Gross 2002; Myers-Scotton, 1997; Myers-Scotton & Jake, 2000).

In the generativist approach, scholars claimed that the underlying grammar of the languages involved were what constrained CS rather than the surface constraints. For example, Woolford (1983) argued that in a CS utterance each grammar contributes part of the sentence (cf. Belazi, Rubin & Toribio, 1994; Di Sciullo et al., 1986). Similarly, MacSwan (1999) argued that the constraints accounting for monolingual grammars, described in the Minimalist Program (Chomsky, 1995), should also account for CS/bilingual grammars.

A psycholinguistically inspired approach, the MLF (Myers-Scotton, 1993), instead assumes an asymmetry between a matrix language providing the morphosyntactic frame (i.e. the functional elements), and an embedded language providing lexical elements.

CS in the nominal domain

Observations of switches within NPs abound in the literature. Timm (1975) and Lipski (1978) both reported frequent mixed NPs in Spanish–English bilingual production especially between determiners and nouns (DetN). Pfaff (1979), noting similar patterns, suggested that such switches occur

since no structural conflicts arise between the two languages, a notion further developed by Poplack (1980) in the *equivalence constraint*.

Other observations suggested a switching asymmetry. Joshi (1985) noted that in Marathi–English combining Marathi determiners with English nouns was acceptable (*kati chairs*), but not the inverse (**the khurcya*). He introduced the notion of a Matrix language and proposed the *asymmetry constraint*, stating that 'switching a category of the matrix language to a category of the embedded grammar is permitted, but not vice-versa' (Joshi, 1985, p. 192).

In the same vein, the MLF approach (Jake et al., 2002) proposed the *bilingual NP hypothesis*, stating that determiners in mixed nominal constructions should come from the matrix language of the clause, with nominal constructions in the embedded language being permitted but not preferred. Investigating Welsh–English bilinguals, Deuchar (2005, 2006) found that all mixed NPs consisted of a Welsh determiner (Det) and an English noun (N) in utterances where the matrix language was Welsh, thereby lending support to the hypothesis.

Liceras et al. (2008) while examining Spanish–English child and Spanish-dominant adult bilinguals, observed a preference for Spanish determiners with English nouns (e.g. *la chair*) over English determiners with Spanish nouns (e.g. *the silla*). They argued that the findings supported a generativist view where the language with the richest array of 'uninterpretable *phi* features' provides the surface realization of the functional category. Since Spanish determiners carry two uninterpretable features (gender and number), they will be dominant.

A few studies directly testing the MLF (determiners from the matrix language) against generativist predictions (determiners from the language with most *phi* features) in Spanish–English and Welsh–English bilinguals have found results to be either broadly consistent with both sets of predictions or inconclusive (Fairchild and van Hell, 2015; Herring, Deuchar, Parafita Couto, & Moro Quintanilla, 2010). Herring et al. (2010) found that the generativist account was successful in explaining all their Welsh–English data and most of the Spanish–English data. However, they also observed that the success of the generativist account was due to the fact that the language of the verb was almost always Welsh or Spanish, i.e. languages with grammatical gender. In the few clauses where the finite verb was in English, an English determiner was usually found contrary to generativist predictions. Fairchild and van Hell (2017) experimentally examined determiner–noun switches in Spanish–English bilinguals. Their results did not support the predictions of either model. However, they discuss two factors that may have affected their results: (a) they were focusing on externally-induced switches rather than spontaneous, natural switches; and (b) their participants were English-dominant.

More complex NPs have also been examined. Di Sciullo et al. (1986) found noun–adjective switches in Italian–English NPs despite different preferred adjective–noun word orders in these languages. Cantone and MacSwan (2009), investigating German–Italian NPs, concluded that the language of the adjective determined word order. Vanden Wyngaerd (2016) found support for this generalization in French–Dutch NPs. In contrast, Parafita Couto, Deuchar and Fusser (2015), when examining English–Welsh, found that the MLF prediction (where the matrix language determines adjective position) better accounted for the data.

In sum, despite the comparative wealth of observations and theoretical accounts offered to explain them, the evidence is still contradictory. To assess whether this is an artefact of different methods and coding schemes (cf. Pérez-Leroux, O'Rourke & Sunderman, 2014), we must compare similar data types from several language pairs coded identically.

The current study

The current study tests predictions derived from generativist accounts (Cantone & MacSwan, 2009; Liceras et al., 2008) and the MLF account (Myers-Scotton, 1993, 2002) concerning the mechanisms

Table 1. Predictions.

Framework	Prediction
Generativism	**Determiner:** The language with most *phi* features provides the determiner (i.e. Welsh, Spanish, Dutch).
	Word order: The language of the adjective sets word order.
MLF	**Determiner:** The Matrix Language provides the determiner (i.e. Welsh, Spanish, Papiamento).
	Word order: The Matrix Language sets word order (i.e. Welsh, Spanish, Papiamento).

underpinning NP-internal switches. We extend previous work by focusing on both simplex and complex NPs, specifically the language of the determiner and noun–adjective word order.

For the language of the determiner, generativism predicts that the determiner is provided by the language with more grammaticized/*phi* features. The MLF account instead predicts that the determiner is provided by the matrix language and determined by finite verb morphology.

For noun–adjective word order, the prediction derived from Cantone and MacSwan (2009) is that the language of the adjective determines its position. The MLF predicts that adjectives occur in the position matching the matrix language.

We test these predictions on three language pairs: Welsh–English, Spanish–English, and Papiamento–Dutch. These languages differ in gender and noun–adjective word order: Spanish, Welsh and Dutch have gender, English and Papiamento do not; Spanish, Welsh and Papiamento prefer post-nominal adjectives, Dutch and English prefer prenominal ones.

Table 1 summarises the predictions from the two accounts regarding determiners and adjective–noun order.

Method

Corpora

We draw on three publicly available corpora (see Appendix 1).

Welsh–English. The Welsh–English corpus (Deuchar, Davies, Herring, Parafita Couto & Carter, 2014) consists of 40 hours of recordings, with 151 speakers from various educational backgrounds (81 female, age range 10–80 years) engaged in dyadic conversation on a free topic. Of the speakers, 63% ($n = 95$) assessed their own proficiency as equally high in both languages. Of the remaining 56 speakers, 66% ($n = 37$) reported being more proficient in Welsh. The recordings have been transcribed in CHAT format (see MacWhinney, 2000). An example of a nominal switch is found in (1) (English in bold):

(1) *yr # dynes* **crazy**

DET woman crazy

'…that crazy woman' (Stammers 5:512)

Spanish–English. The Spanish–English corpus (Deuchar et al., 2014) – collected in Miami, Florida (FL) – consists of 30 hours of recordings with 85 speakers (52 female, age range 11–78 years). Of

the speakers, 62% (n = 53) assessed their own proficiency to be equally high in both languages. Of the remaining 32 speakers, 69% (n = 22) reported being more proficient in English. The corpus was collected and transcribed in the same fashion as the Welsh–English corpus. An example of a nominal switch is found in (2) (English in bold):

(2) *esto es un pequeño* **pocket** …

DEM is a small pocket

'This is a small pocket…'

[Herring 63, *KEV]

Papiamento–Dutch. The Papiamento–Dutch corpus consists of three hours of free conversation from six four-party conversations involving 25 early functional Papiamento–Dutch bilinguals (15 female, age range 18–61) born in Aruba (n = 10), Curaçao (n = 9) and Surinam (n = 1), but resident in the Netherlands at the time of recording (Gullberg, Indefrey, & Muysken, 2009). Their educational background ranged from vocational training to university education. All participants reported using both languages to the same extent daily in a range of situations and to habitually code-switch with other bilinguals. Nevertheless, 24 out of the 25 speakers reported that Papiamento was their 'best language'.

The conversations are transcribed using standard Dutch and Aruban orthography with phonetic modifications, hesitations and overlapping speech marked. The transcripts are glossed and tagged for language and word class using the coding scheme from Muysken, Kook, and Vedder (1996). An example of a nominal switch is found in (3) (Papiamento in bold).

(3) ***un*** *elftal* ***mixto***

det.INDEF eleven mixed

'**a mixed** eleven [football team]' (14, 0141)

Data and coding

The current analyses are based on subsets of the Welsh–English (42 speakers) and Spanish–English corpora (19 speakers), and on the entire Papiamento–Dutch corpus (25 speakers). Table 2 specifies the details. The subsets match the bigger corpora with regard to speakers' gender and age.

For the purposes of these analyses, we re-coded all three datasets identically. Each NP was coded for the grammatical category of the constituting parts (e.g. determiner–adjective–noun, determiner–noun, etc.). 'Determiners' included articles, demonstrative and possessive pronouns. 'Adjectives' (A) included adjectives and adjectivised nouns.

We also coded each NP-internal grammatical element for language. Each word was tagged for source language (Spanish, Welsh, English, Papiamento or Dutch). 'Foreign' words found in the lexicon of monolingual speakers and in dictionaries were tagged as a separate category (e.g. *internet)* unless the phonetic realization determined source language.

From the three re-coded corpora we extracted all monolingual and mixed simplex NPs (DetN) and complex NPs with determiners (determiner–adjective–noun; DetAN/NA) (see Appendix 2). We excluded all cases of mixed NPs that included words where language was ambiguous (WE n = 12; SE n = 6; PD n = 10).

Table 2. The three corpora.

Welsh–English	Spanish–English	Dutch–Papiamento
• 151 speakers (81 female) • dyads • 18:40 h/40 (42 speakers) selected • Collected at the Centre for Research on Bilingualism, Bangor, UK • Mixed NPs *n*=171	• 85 speakers (52 female) • dyads • 5:27 h/20 (19 speakers) selected • Collected in Miami, FL, USA • Mixed NPs *n*=98	• 25 speakers (15 female) • four-party conversations • 3 h (25 speakers) selected • Collected at the MPI for Psycholinguistics, the Netherlands • Mixed NPs *n*=60

Table 3. Distribution of monolingual and mixed NPs in the three corpora (W = Welsh; E = English; S = Spanish; P = Papiamento; D = Dutch). Simplex = DetN; Complex = DetNA/AN.

	Monolingual		Mixed		Total
	DetN	DetNA/AN	DetN	DetNA/AN	
W–E	W 2932 E 126	W 426 E 18	146	25	3673
S–E	S 1210 E 1128	S 144 E 401	92	6	2981
P–D	P 989 D 66	P 68 D 18	41	19	1201

We examined the surrounding clause to determine the matrix language based on the finite verb (following Herring et al., 2010). The matrix language was exclusively Welsh and Papiamento, respectively, in the Welsh–English and Papiamento–Dutch corpora. In the Spanish–English corpus, Spanish was the matrix language in 79% of the cases (79/100) and English in 10% (10/100). There were 11 cases where the matrix language could not be identified.

In the discussions below, we shall refer to A-languages (Welsh, Spanish and Papiamento) and B-languages (English and Dutch, in bold font).

Analyses

We analysed the data using a linear regression model using the lm command in R statistical software (version 0.98.953) to examine distribution of A- and B-languages across word class and word order in the corpora.

Results

Table 3 presents the distribution of monolingual and mixed NPs in the corpora. Monolingual NPs are presented for descriptive purposes but the analyses focus only on the mixed NPs (Welsh–English $n = 171$; Spanish–English $n = 98$; Papiamento–Dutch $n = 60$).

Table 3 also shows the distribution of mixed simplex NPs (DetN) versus mixed complex NPs (DetNA/AN). The patterns are remarkably similar. In all corpora the majority of the mixed NPs are of the simplex type (85% for Welsh–English, 94% for Spanish–English and 68% for Papiamento–Dutch, respectively).

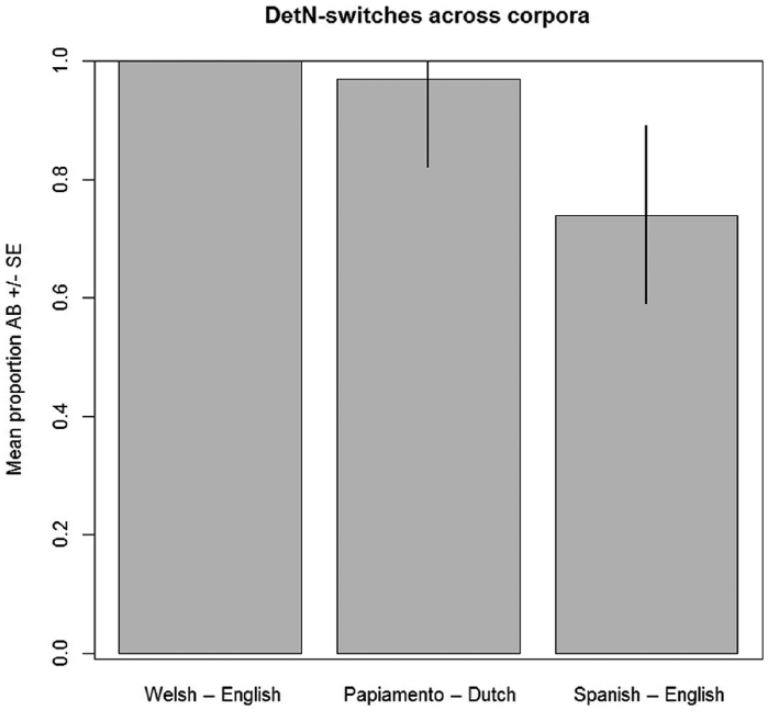

Figure 1. DetN switches across corpora.

Mixed simplex NPs

First, we investigated the distribution of languages across Dets and Ns in DetN switches. Specifically we examined the extent to which Dets came from A-languages (Welsh, Spanish and Papiamento) and Ns from B-languages (English and Dutch, in bold), as shown in examples (4)–(6). The data were aggregated by subject and language combination, and proportions of AB-switches were calculated for each row of the aggregated dataset. A linear regression model was fit to these proportions with corpus as fixed effect. Since the AB-distribution pattern manifested itself most strongly in the Welsh–English corpus, this corpus was coded as the baseline against which the other corpora were compared. The intercept of the linear model therefore represents the Welsh–English corpus. Figure 1 shows the distribution of DetN switches over languages in mixed simplex NPs.

(4) *y* ***thing*** (Welsh–English, Davies 5, Line 970)
Det^{W} thing^{E}
'the thing'

(5) *el* ***environment*** (Spanish–English, Herring85, *JAD)
Det^{S} environment^{E}
'the environment'

(6) *e* ***voetganger*** (Papiamento–Dutch, 03, 32)
Det^{P} pedestrian^{D}
'the pedestrian'

Table 4. Coefficients for the linear regression model (AB-switches vs BA-switches in bare DetN combinations across corpora).

	Estimate	*SE*	*T*	*p*
(Incercept)	1.00000	0.13440	7.441	$104\times^{-08}$
langcombPD	–0.03125	0.15397	–0.203	0.8403
langcombSE	–0.26062	0.15289	–0.705	0.0971

Table 5. Mixed complex NPs and distribution of language, word class and word order (A = Welsh, Spanish, Papiamento; B = English, Dutch).

	ABB		AAB		ABA	
	DetAN	DetNA	DetAN	DetNA	DetAN	DetNA
W–E	56% (14/25)	0	8% (2/25)	4% (1/25)	8% (2/25)	24% (6/25)
S–E	67% (4/6)	0	17% (1/6)	0	0	17% (1/6)
P–D	58% (11/19)	0	16% (3/19)	11% (2/19)	0	16% (3/19)

Figure 1 reveals that in all corpora the AB-switches are the most frequent (100% in Welsh–English; 97% in Papiamento–Dutch; 74% in Spanish–English). AB-switches in the Spanish–English corpus are numerically fewer than in the other corpora but the difference does not reach significance (see Table 4).

Overwhelmingly, Dets also occur in the same language as the matrix language of the clause (100% match in Welsh–English and Papiamento–Dutch with two exceptions in Spanish–English).

Mixed complex NPs

Next, we examined the distribution of Det, N and A in all mixed complex NPs across languages and word order.

Table 5 reveals that the most frequent type of complex mixed NP in all corpora are DetAN with ABB language distributions, that is, combinations of Dets from A-languages (Welsh, Papiamento and Spanish) and AN-clusters from B-languages (English and Dutch) (examples (7)–(9)). The data points are too few for a statistical comparison to be made across the corpora but the pattern is quite clear. The frequency of other combinations is very low (range 0–6).

(7) *y* ***Belgian loaf*** (Welsh–English, Robert 5, Line 150)

Det^{W} Belgian^{E} loaf^{E}

‘the Belgian loaf”

(8) *los* ***dry walls*** (Spanish–English, Herring59, *KEV)

Det^{S} dry^{E} walls^{E}

‘ the dry walls’

(9) *un* ***moeilijke keuze*** (Papiamento–Dutch, 16, 0369)

Det^{P} difficult^{D} choice^{D}

'a difficult choice'

Discussion

The analyses of the mixed NPs in the three corpora reveal three key patterns;

1. Simplex switches between Det and N dominate in all language pairs;
2. In mixed NPs Dets overwhelmingly come from Welsh (+gender), Spanish (+gender) and Papiamento (-gender), respectively;
3. Preposed adjectives are the most common in all language pairs followed by Ns in the same language.

The determiner results provide evidence against generativist predictions (which were Dets in Welsh, Spanish and Dutch) since Papiamento provided Dets. Critically, Papiamento should not provide Dets according to the generativist prediction as it lacks gender/*phi* features. Instead, we can see the matrix languages (Welsh, Spanish, English and Papiamento) provided Dets (the results show an overwhelming match between Dets and matrix languages).

In contrast, the word order results support predictions from both approaches. Dets in Welsh, Spanish and Papiamento are followed by AN clusters in English and Dutch with As in the prenominal position as is typical of these languages. These results match the generativist predictions. Arguably, they also match MLF predictions but in a different form than posited in Table 1. In examples (6)–(9) As and Ns come from the same language, and critically *not* the matrix (A-) language of the clause but from the embedded (B-) languages. However, the MLF allows for these constructions referring to such AN clusters as 'embedded language islands' (Myers-Scotton, 1993). In such islands, the grammar of the embedded language prevails.

Generally, it is striking that switches predominantly occur between Dets and AN clusters – not between As and Ns. Moreover in the nine instances of switches between A and N found, the A position always matches the matrix language in accordance with MLF but not necessarily with generativist predictions. Counterexamples to generativist predictions were found in the Welsh–English (n = 1) and the Papiamento–Dutch corpus (n = 6) where English and Dutch As occurred postnominally.

Arguably then, the MLF predictions best fit the results overall whereas the generativist predictions are mainly supported for word order but not for gender instantiated on Dets. These findings highlight the importance of comparing multiple language pairs with similar coding. Two of the corpora did not allow any distinction to be made between the theoretical accounts; only the Papiamento–Dutch corpus did.

All three corpora display clear asymmetries in the switching patterns between A- and B-languages. A pertinent question is what determines the direction of those asymmetries. An extralinguistic factor – language dominance – is often discussed as a possible candidate (Liceras, Fernández Fuertes, & Klassen, 2016) either at the community level (Carter, Deuchar, Davies, & Parafita Couto, 2011; Parafita Couto, Davies, Carter, & Deuchar, 2014) or at the individual level (Pérez-Leroux et al., 2014). The majority of the Welsh–English and Papiamento–Dutch bilinguals reported being more dominant in Welsh and Papiamento, respectively. In both cases, the CS patterns seem to fit with individual and community-wide dominance. In contrast, in the

Spanish–English corpus, individual dominance varies more. This may explain why we find more cases of English Dets and Spanish N than expected (26%, cf. Figure 1). Yet, despite this variation, the CS patterns are similar to the other corpora. Self-reported dominance may therefore not be the best predictor for the patterns observed.

Another possibility is that switching directionality is affected by another extra-linguistic factor. Blokzijl, Deuchar & Parafita Couto (2017) has recently suggested that switches tend to be towards the language of power or the language with superior social status. The patterns observed in the three corpora with Dets from what could be described as the minority languages (Welsh in the UK; Spanish in Miami; Papiamento in the Netherlands) and lexical material from 'majority' languages (English in the UK and in Miami; Dutch in the Netherlands) are certainly consistent with this suggestion.

A related but slightly different construct is frequency of use. Psycholinguistically, frequency effects are known to affect ease of processing in comprehension and production (Ellis, 2002 for an overview). The more frequent a word combination is in the input, the more likely the words are to be named together again. Dets are among the most frequent items in a language, and if Dets are prevalent in one language in a bilingual's input, this should reinforce the same language choice in production. Such an exposure-driven account is posited by Valdés, Kroff (2016) suggesting that bilingual speakers converge on conventional production patterns. The 'islands' in the mixed complex NPs – accounting for the prenominal adjective word order advantage – may at least in part be determined by speakers' relative experience with the adjective–noun combinations in question. These issues should be studied experimentally, drawing inspiration from recent studies of multiword units and collocations (e.g. Gyllstad & Wolter, 2016; Sprenger 2003). The challenge will be to find evidence for frequency patterns in bilingual corpora and even more so to determine with what input frequencies a bilingual individual operates (cf. Green, 2011).

Other issues to explore include priming, a cousin of the frequency effect, whereby what has just been processed is likely to be repeated (e.g. Bock, 1986). One speaker's CS seems to facilitate another speaker's similar switching (Fricke & Kootstra, 2016; Kootstra, van Hell & Dijkstra, 2010), which suggests that this is an important mechanism to consider.

To conclude, the current study has provided new empirical data on nominal CS from language pairs not previously considered. It has revealed robust patterns across three corpora and taken a step towards disentangling two theoretical accounts. Obviously, many issues remain to be explored. Clearly, further descriptive and experimental work is needed – as are accessible corpora – if we are to understand the complex phenomenon that is CS.

Acknowledgements

We gratefully acknowledge funding and support from the British Council Researcher Exchange Programme; from the Max Planck Institute for Psycholinguistics, Nijmegen; and from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO; MPI 56-384, *The Dynamics of Multilingual Processing*, M. Gullberg, P. Indefrey, W. Klein). We also thank the following people for their help with data collection, transcription, coding, and other data treatment; in Bangor: F. Aveledo, D. Carter, P. Davies, M. Deuchar, K. Donnelly, M. Fusser, J. Herring, S. Lloyd-Williams, E. Robert, and J. Stammers; in Nijmegen: I. Brasileira, A. van Doorn, Z. Fransisco, F. Hellwig, N. Josephina, J. Lourens, B. du Pau, M. del Rosario, and M. Williams; in Lund: Henrik Garde, Josine Greidanus. Finally, we are grateful to P. Indefrey and P. Muysken for their input to the Papiamento-Dutch project.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Azuma, S. (1993). The frame-content hypothesis in speech production: Evidence from intrasentential code switching. *Linguistics*, *31*, 1071–1093.

Belazi, H., Rubin, E., & Toribio, A. J. (1994). Code-switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, *25*, 221–237.

Bentahila, A., & Davies, E. E. (1983). The syntax of Arabic–French code-switching. *Lingua*, *59*(4), 301–330.

Berk-Seligson, S. (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society*, *15*(3), 313–348.

Blokzijl, J., Deuchar, M., & Parafita Couto, M. C. (forthcoming, 2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bullock, B., & Toribio, A. J. (Eds.) (2009). *The handbook of linguistic code-switching*. Cambridge: Cambridge University Press.

Cantone, K. F., & MacSwan, J. (2009). The syntax of DP-internal codeswitching. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to codeswitching* (pp. 243–278). Amsterdam: John Benjamins Publishing.

Cantone, K. F., & Müller, N. (2008). Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*, *118*, 810–826.

Carter, D., Deuchar, M., Davies, P., & Parafita Couto, M. (2011). A systematic comparison of factors affecting the choice of matrix language in three bilingual communities. *Journal of Language Contact*, *4*, 1–31.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Deuchar, M. (2005). Congruence and code-switching in Welsh. *Bilingualism: Language and Cognition*, *8*(3), 1–15.

Deuchar, M. (2006). Welsh–English code-switching and the Matrix Language Frame model. *Lingua*, *116*(11), 1986–2011.

Deuchar, M. (2012). Code-switching. In C. A. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 675–664). New York: Wiley.

Deuchar, M., Davies, P., Herring, J. R., Parafita Couto, M., & Carter, D. (2014). Building bilingual corpora: Welsh–English, Spanish–English and Spanish–Welsh. In E. M. Thomas & I. Mennen (Eds), *Advances in the study of bilingualism* (pp. 93–110). Bristol: Multilingual Matters.

Di Sciullo, A. M., Muysken, P., & Singh, R. (1986). Government and code-switching. *Journal of Linguistics*, *22*, 1–24.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188.

Fairchild, S., & van Hell, J. G. (2017). Determiner–noun code-switching in Spanish–English heritage speakers. *Bilingualism: Language and Cognition*, *20*(1), 150–161.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201.

Gordon, R. G., Jr. (Ed.) (2005). *Ethnologue: Languages of the world* (15th ed.). Dallas, TX: SIL International.

Green, D. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology*, *2*(103). doi:10.3389/fpsyg.2011.00103.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). In Bullock, B. E., & Toribio, A. J. (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge: Cambridge University Press.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*, 296–323.

Herring, J., Deuchar, M., Parafita Couto, M. C., & Moro Quintanilla, M. (2010). 'When I went to Canada, I saw the madre': Evaluating two theories' predictions about codeswitching between determiners and nouns using Spanish–English and Welsh–English bilingual corpora. *International Journal of Bilingual Education and Bilingualism*, *13*, 553–573.

Jake, J. L., Myers-Scotton, C., & Gross, S. (2002). Making a minimalist to codeswitching work: Adding the Matrix Language. *Bilingualism, Language & Cognition*, *5*, 69–91.

Joshi, A. (1985). Processing of sentences with intrasentential code-switching. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 190–204). Cambridge: Cambridge University Press.

Kootstra, G. J., van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language*, *63*, 210–231.

Kouwenberg, S., & Murray, E. (1994). *Papiamentu*. München: Lincom Europa.

Liceras, J., Fernández Fuertes, R., & Klassen, R. (2016). Language dominance and language nativeness: The view from English–Spanish codeswitching. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish–English codeswitching in the Caribbean and the US* (pp. 107–138). Amsterdam: John Benjamins.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional–lexical mixings. *Lingua*, *118*, 827–851.

Lipski, J. (1978). Code-switching and the problem of bilingual competence. In M. Paradis (Ed.), *Aspects of bilingualism* (pp. 250–264). Columbia: Hornbeam Press.

MacSwan, J. (1999). *A minimalist approach to intra-sentential code-switching*. New York: Garland.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Muysken, P., Kook, H., & Vedder, P. (1996). Papiamento/Dutch code-switching in bilingual parent-child reading. *Applied Psycholinguistics*, *17*, 485–505.

Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Oxford University Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, C., & Jake, J. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, *38*, 1053–1100.

Parafita Couto, M. C., Davies, P., Carter, D., & Deuchar, M. (2014). Factors affecting code-switching. In E. Thomas & I. Mennen (Eds.), *Unraveling bilingualism: A cross-disciplinary perspective* (pp. 111–140). Bristol: Multilingual Matters.

Parafita Couto, M. C., Deuchar, M., & Fusser, M. (2015). How do Welsh–English bilinguals deal with conflict? Adjective–noun order resolution. In G. Stell & K. Yakpo (Eds.), *Code-switching between structural and sociolinguistic perspectives* (pp. 64–84). Berlin: De Gruyter.

Pérez-Leroux, A. T., O'Rourke, E., & Sunderman, G. (2014). Language dominance and codeswitching asymmetries. In J. McSwan (Ed.), *Grammatical theory and bilingual codeswitching* (pp. 283–312). Cambridge, MA: MIT Press.

Pfaff, C. W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, *55*(2), 291–318.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en Español*: Toward a typology of code-switching. *Linguistics*, *18*, 581–618.

Poplack, S. (1981). Syntactic structure and social function of code-switching. In R. Duran (Ed.), *Latino language and communicative behavior* (pp. 169–184). Norwood, NJ: Ablex.

Sprenger, S. (2003). *Fixed expressions and the production of idioms*. (Unpublished PhD dissertation). Max Planck Institute for Psycholinguistics, Radboud University.

Timm, L. A. (1975). Spanish–English code-switching: El porqué y how-not-to. *Romance Philology*, *28*(4), 473–482.

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish–English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. Guzzardo Tamargo, C. Mazak, & M. C. Parafita Couto (Eds.), *Spanish–English codeswitching in the Caribbean and the US* (pp. 281–300). Amsterdam: John Benjamins.

Vanden Wyngaerd, E. (2017). The adjective in Dutch–French codeswitching: Word order and agreement. *International Journal of Bilingualism*, *21*(4), 454–473. DOI: 10.1177/1367006916632302.

Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry*, *14*, 520–536.

Author biographies

Maria Carmen Parafita Couto is university lecturer at the Leiden University Center for Linguistics. She studies the impact of bilingualism and language contact on language structure, mostly the syntax of code-switching.

Marianne Gullberg is full professor of Linguistics at Lund University, and chair/director of Lund University Humanities Lab, an independent department for research technology and method. Her research focuses on acquisition and real time language use in adult second language users or bi-/multilingual speakers. She also studies the bimodal nature of language taking into account gesture production and comprehension in acquisition and bilingualism.

Appendix 1

URLs to the corpora:
Welsh–English
Retrieved from http://bangortalk.org.uk/speakers.php?c=siarad
Spanish–English
Retrieved from http://bangortalk.org.uk/speakers.php?c=miami
Papiamento–Dutch
Retrieved from http://www.mpi.nl/resources/data/browsable-corpora-at-mpi

Appendix 2: Mixed NPS

171 Welsh–English mixed NPs

Davies14,Line709,(y)rAnglican chaplaincy
Davies 4,Line 184,y gay city [of the world]
Davies 4,Line 704,(y)r bloody fridge
Davies 5,Line 46,yr external hardrive
fusser 11,Line 11,y timing belt
fusser 26,Line 36,yr general office
Lloyd 1,Line 165,yr Greek yogurt
Robert 3,Line 265,y flowering cherry
Robert 3,Line 271,y belgian loaf
Robert 5,Line 3,yr african violet
Robert 5,Line 150,y belgian loaf
Robert 3,Line 1225,y sealed bids
Robert 3,Line 562,y nursing training
fusser 3,Line 38,yr amazing screeching car
fusser 11,Line 1086,yr hen gomputers
fusser 26,Line 376,(y)r hên cowpox
Robert 3,Line 389,y raw peth
Robert 3,Line 368,(y)r blinking speakers mawr
fusser 17,Line 328,y pethau glacial
Davies 5,Line 1251,dy lessons dreifio
fusser 11,Line 182,yr mileage bychan
Robert 1,Line 719,y spear_tackle enwog
Stammers 6,Line 530,y bid cyntaf
Stammers 6,Line 1277,(y)r hail mawr

Stammers 6,Line 640,eu budgets flwyddyn yma
Davies 1,Line 580,yr eighteenths
Davies 1,Line 899,y blood tests
Davies 11,Line 1036,y Wirral
Davies 3,Line 142,y surround_sound
Davies 3,Line 231,(y)r thing
Davies 3,Line 339,y swimming_pool
Davies 3,Line 346,y swimming_pool
Davies 3,Line 515,y commitment
Davies 3,Line 696,yr subwoofers
Davies 3,Line 708,y thing
Davies 3,Line 741,y speakers
Davies 3,Line 843,(y)r laptop
Davies 4,Line 139,yr destination
Davies 4,Line 273,y secretary
Davies 4,Line 407,yr equator
Davies 4,Line 411,(y)r foyer
Davies 4,Line 702,(y)r fridge
Davies 4,Line 703,(y)r fridge
Davies 4,Line 927,yr head waiter
Davies 4,Line 960,yr head waiter
Davies 4,Line 1157,(y)r moisture
Davies 4,Line 1178,yr twenty_fifth
Davies 4,Line 1210,y message
Davies 4,Line 1213,(y)r agencies
Davies 5,Line 308,yr universities
Davies 5,Line 763,yr Commonwealth
Davies 5,Line 781,y World_Cup
Davies 5,Line 849,y fix
Davies 5,Line 970,y thing
Davies 5,Line 89,(y)r Sky movie channels
Davies 5,Line 104,(y)r Cartoon-Network channels
Davies 5,Line 205,fy degree
Davies 5,Line 296,yr injection
Davies 5,Line 449,y cathedral
Davies 5,Line 724,y photography
Davies 5,Line 946,y loan
Davies 9,Line 205,y keyboard
fusser 17,Line 1353,yr amser amser exposure
fusser 11,Line 200,yr average
fusser 11,Line 307,yr Japanese
fusser 11,Line 383,(y)r book value
fusser 11,Line 387,y fore_courts
fusser 11,Line 1385,yr breeding
fusser 13,Line 617,y forecast
fusser 17,Line 273,y reflection
fusser 17,Line 312,y ropeway
fusser 17,Line 312,y peiriannau ropeway

fusser 17,Line 394,y ridge
fusser 17,Line 471,y composition
fusser 17,Line 1010,y composition
fusser 17,Line 1251,yr angle
fusser 17,Line 1308,y reflection
fusser 17,Line 1485,y geology
fusser 26,Line 92,yr sister
fusser 26,Line 95,y sister
fusser 26,Line 148,y National Service
fusser 26,Line 330,yr twins
fusser 26,Line 343,yr fever
fusser 26,Line 431,yr disease
fusser 26,Line 448,y structure
fusser 26,Line 481,(y)r dresser
fusser 26,Line 494,(y)r clearings
fusser 26,Line 587,(y)r computer
fusser 26,Line 761,y celandine
fusser 26,Line 764,(y)r dail celandine
fusser 27,Line 32,y country
fusser 27,Line 149,y menthol stuff
fusser 27,Line 389,(y)r volcanoes
fusser 27,Line 728,y belt attachment
fusser 3,Line 95,yr thing
fusser 3,Line 648,y microwave
fusser 30,Line 313,y spray tan
fusser 30,Line 659,y vows
fusser 5,Line 717,(y)r laptop
fusser 5,Line 738,dy laptop
fusser 5,Line 741,(y)r laptop
fusser 5,Line 750,y laptop
fusser 5,Line 806,y laptop
fusser 5,Line 916,yr strips arian
fusser 5,Line 753,eich laptop
fusser 4,Line 86,y crater
fusser 4,Line 87,y strays
fusser 4,Line 507,y fireworks
Lloyd 1,Line 84,yr date
Lloyd 1,Line 338,yr machine
Lloyd 1,Line 518,y rescue team
Lloyd 1,Line 588,y wash
Lloyd 1,Line 591,y wash
Robert 1,Line 32,y registrar
Robert 1,Line 151,(y)r tea_lights
Robert 1,Line 154,(y)r caterer
Robert 1,Line 195,(y) booking
Robert 1,Line 202,(y)r truffles
Robert 1,Line 203,(y)r tea_lights
Robert 1,Line 203,(y)r truffles

Robert 1,Line 206,y truffles
Robert 1,Line 328,y Coal_Exchange
Robert 1,Line 374,y Coal_Exchange
Robert 1,Line 673,y thing
Robert 1,Line 712,y thing
Robert 2,Line 66,(y)r internet
Robert 2,Line 317,y safety pins
Robert 2,Line 666,(y)r pharmacists
Robert 2,Line 913,yr off_licence
Robert 3,Line 499,fy day off
Robert 3,Line 581,y training
Robert 3,Line 594,y midwifery
Robert 3,Line 1034,y flannel
Robert 5,Line 13,y florists
Robert 5,Line 160,yr ground_elder
Robert 6,Line 50,y rail
Robert 6,Line 387,yr hand luggage
Robert 6,Line 479,yr bridesmaids dresses
Robert 6,Line 741,y Russian
Robert 6,Line 745,yr Russian
Robert 7,Line 472,y web
Robert 7,Line 475,y web
Robert 7,Line 501,y website
Robert 7,Line 527,y mileage
Robert 7,Line 747,y phase two
Robert 7,Line 1198,(y)r two thousand and five
Robert 1,Line 213,yr box_office
Robert 1,Line 267,y mannerisms
Robert 1,Line 329,(e)i goatee beard
Robert 1,Line 102,y chemistry
Robert 1,Line 316,(y)r love God
Robert 3,line 1104,(y)r mortgage
Robert 3,Line 36,y Welsh_Cup
Robert 3,Line 139,(y)r storylines
Robert 3,Line 358,(y)r health_and_safety
Robert 3,Line 458,(y)r infection
Robert 3,Line 594,yr wing mirror
Robert 3,Line 863,y judge
Robert 3,Line 1247,y seal
Robert 3,Line 1287,(y)r electrician
Robert 6,Line 588,yr Corns
Robert 3,Line 485,y nebulizer
Stammers 1,Line 490,y fairway
Stammers 1,Line 820,y consultant
Stammers 2,Line 121,yr World_Cup
Stammers 2,Line 196,y breeze fairy
Stammers 2,Line 392,yr waste
Stammers 2,Line 810,yr gel

Stammers 6,Line 391,dy fowels
Stammers 6,Line 1244,(y)r boiler
Stammers 9,Line 117,y niece

98 Spanish–English mixed NPs

File, speaker, NP
Herring49, *SEB:, la cheerleader pesada
Herring4, *PAI:, un healthy store
Herring50, *RIC:, el bad guy
Herring59, *KEV:, los dry walls
Herring66, *KEV:, otro zip code
Herring63, *KEV:, un pequeño pocket
Herring1,*LAU:,the madre
Herring2,*SAR:,un special
Herring3,*SAR:,el food festival ese
Herring5,*PAI:,el website
Herring6,*SAR:,el trailer
Herring7,*SAR:,un website
Herring8,*SAR:,el countdown
Herring9,*PAI:,el countdown
Herring10,*SAR:,el countdown
Herring11,*SAR:,un site
Herring12,*MIG:,un mall
Herring13,*MIG:,qué mall
Herring14,*MIG:,un lounge
Herring15,*TIM:,del ATM
Herring16,*MIG:,el ATM
Herring17,*MIG:,un ATM
Herring18,*MIG:,al ATM
Herring19,*MIG:,mi ID
Herring20,*MIG:,el ATM
Herring21,*TIM:,el ATM
Herring22,*TIM:,un fee
Herring23,*MIG:,un jeans
Herring24,*TOM:,el spring break
Herring25,*TOM:,al mall
Herring26,*TOM:,un house party
Herring27,*TOM:,el summer
Herring28,*MIG:,su wall
Herring29,*MIG:,mi profile
Herring31,*MIG:,el expressway
Herring32,*ASH:,los brochures
Herring33,*ASH:,mi printer
Herring34,*ASH:,el payroll
Herring35,*ASH:,el sales office
Herring36,*ASH:,el orientation
Herring37,*ASH:,ese internship

Herring38,*ASH:,el internship
Herring39,*ASH:,un intern
Herring40,*JAC:,un restaurant
Herring41,*JAC:,un hotdog
Herring42,*ASH:,un hotdog
Herring43,*ASH:,al valet
Herring45,*SEB:,un shot
Herring46,*SEB:,un shot
Herring47,*SEB:,that enciclopedia
Herring48,*SEB:,el profile
Herring51,*CLA:,the manguera
Herring52,*CLA:,the manguera
Herring53,*CLA:,the manguera
Herring54,*LUK:,the amo
Herring55,*CLA:,a vieja
Herring56,*SOF:,los speed bumps
Herring57,*KEV:,todos los leftovers
Herring58,*SOF:,un hammock
Herring60,*KEV:,un camper
Herring61,*KEV:,un RV
Herring62,*SOF:,that palmero
Herring64,*KEV:,el town of Miami Lakes
Herring65,*KEV:,del post office
Herring67,*KEV:,ese zip code
Herring68,*SOF:,el W1040
Herring69,*SOF:,la 1040
Herring70,*KEV:,las 1040
Herring71,*SOF:,el IRS
Herring72,*KEV:,al IRS
Herring73,*KEV:,el 1099
Herring74,*KEV:,un barbecue place
Herring75,*KEV:,el spring break
Herring76,*KEV:,un MP3 player
Herring77,*KEV:,un recording
Herring78,*SOF:,el sticker
Herring79,*KEV:,un rebuilt
Herring80,*SOF:,la 441
Herring81,*KEV:,el water management district
Herring82,*KEV:,los lily pads
Herring83,*KEV:,un CD
Herring84,*KEV:,el desktop
Herring85,*JAD:,el environment
Herring86,*JAD:,el environment
Herring87,*JUL:,a botánica
Herring88,*CAR:,un tampon
Herring89,*CAR:,un tampon
Herring90,*CAR:,el paperwork
Herring91,*CAR:,ese shift

Herring92,*AME:,un retail store
Herring93,*CAR:,los employees
Herring94,*CAR:,un customer
Herring95,*AME:,al manager
Herring96,*CAR:,el schedule
Herring97,*AME:,el front desk
Herring98,*CAR:,este man
Herring99,*CAR:,los fire extinguishers
Herring100,*HEN:,del manicure

60 Papiamento–Dutch mixed NPs

Speaker, record number, NP
3, 0376, e algemene indruk
2, 0489, un grote teleurstelling
1, 0531, un lange vakantie
16, 0369, un moeilijke keuze
16, 0477, e harde kern
19, 0971, e macho mentaliteit
20, 1056, e gevaarlijkste stad
24, 0225, un vertekende beeld
24, 0230, un totale beeld
22, 0326, un bepaalde drukte
22, 0332, e simpele voetganger
1, 0549, un dushi verblijf
21, 0569, e bendita jointje
15, 0670, un otro divisie
15, 0613, un mucha muhe maagd
18, 0661, un oficina ingenieurs
14, 0141, un elftal mixto
21, 0790, un strijd constante
22, 0514, un voorsprong hopi grandi
16, 0213, nan hoofddoek
1, 0371, un teleurstelling
1, 0372, un teleurstelling
2, 0445, un reden
1, 0499, e reden
3, 0574, un knop
3, 0699, e optocht
10, 0514, un toon
9, 0980, mi buurman
14, 0311, bo doorstroomsysteem
14, 0671, un beeldbuis
14, 0983, un stemrecht
14, 0984, un stemplicht
12, 0995, bo stem
14, 1014, un partij
14, 1017, un kamerbespreking

14, 1118, un regeerakkoord
14, 1129, un regeerakkoord
18, 0076, e werkveld
15, 0228, e geloof
16, 0601, e maatschappij
17, 0933, un vuilnisman
16, 0982, un verblijftehuis
19, 0004, un kadotje
21, 0054, un cultuurschok
21, 0146, e gehakt
20, 0237, un buurjongen
19, 0757, e respect
21, 0790, un strijd
20, 0918, nan mentaliteit
21, 1027, un volksbuurt
22, 0041, e samenleving
24, 0068, e vriendjespolitiek
22, 0272, e gedraging
22, 0332, e voetganger
22, 0344, e bushalte
22, 0349, un kinderwagen
24, 0574, e bevolking
24, 0777, un allochtoon
3, 0373, e eerste
10, 1013, e taal