Word classes in language contact

Yaron Matras, Evangelia Adamou

Word classes in language contact

Yaron Matras and Evangelia Adamou

Abstract

Word classes show different degrees of sensitivity and susceptibility both to contact induced language change (borrowing) and to language mixing in conversation (codeswitching). Our contribution addresses the hierarchical nature of word class arrangement as manifested in language contact settings. We review generalisations that have addressed the question of borrowing constraints on particular categories and examine some of the features that accompany the integration of borrowed items in particular word classes. We then examine some of the insights gained from the analysis of bilingual corpora and their implications for an explanatory model of borrowing and the status of word classes in contact-induced change.

1. Introduction

Since Whitney's (1881) pioneering discussion of the impact of language contact on structural change, studies have addressed the question whether structural categories differ in their susceptibility to borrowing and whether constraints can be identified on borrowability. In this contribution we review key aspects of those discussions with a focus on the distinctive behaviours of different word classes. We begin with a brief survey of generalisations that have addressed the question of borrowing constraints on particular categories. We then examine individual word classes in respect of their contact behaviour and the processes of bilingual speech that initiate contact induced language change. We conclude by addressing the implications for a theoretical and methodological appreciation of word class categorisation.

2. Word classes and borrowing hierarchies

Whitney (1881) posited that word classes react differently to the pressure of language contact. He observed that nouns are more easily borrowed than adjectives and that adjectives in turn are more easily borrowed than verbs. Grammatical function words, he claimed, such as pronouns, articles, prepositions, conjunctions, and numerals, are even less likely to be borrowed and behave in that respect much like grammatical inflection, which tends to be exempted from language mixing (though it may accompany borrowed words, as in the case of the Latin plural ending in the English word *phenomena*). The founders of modern contact linguistics, Einar Haugen and Uriel Weinreich, both commented on the behaviour of categories in contact induced language change.

Haugen (1950: 224-25) refers to a scale of adoptability according to which borrowed nouns are more than three times more numerous than borrowed verbs while adjectives lag behind, followed by adverbs, prepositions and interjections. Pronouns and articles do not appear among the list of borrowings, Haugen emphasises. By way of explanation Haugen proposes that the prevalence of nouns and verbs is connected to vocabulary expansion over a speaker's lifetime (potentially linked to the cultural and technological development of a community) while grammatical items belong to a fixed inventory that is established in early childhood. An important distinction is made between importation of forms, which enriches the vocabulary, and substitution of pre-existing forms; the borrowing of nouns enriches the vocabulary by adding designations for new objects and routines that are specific to certain social and cultural settings. Weinreich (1953: 35) similarly argues that nouns show high borrowability because of their semantic function as designations of new things, the need for affective enrichment or euphemism, the need for differentiation, and a general need for renewal. The prominence of nouns over non-nouns is also noted by Moravcsik (1978), who links borrowability to structural autonomy and semantic transparency; derivational morphology is also observed to be more borrowable than inflectional morphology and unbound elements more borrowable than bound elements. Poplack (2018: 48-50) remarks that the borrowing of nouns compared to other word classes by far exceeds their relative frequency in the receiving language and so the motivation to borrow nouns must be attributed to the richness of semantic content rather than frequency or the structural properties of nouns as potential stand-alone elements.

Building on a discussion by Muysken (1981), Winford (2010: 176) offers the following hierarchy of borrowing by word class: nouns > adjectives > verbs > prepositions > coordinating conjunctions > quantifiers > determiners > free pronouns > clitic pronouns > subordinating conjunctions. Winford notes that this hierarchy has been confirmed by various studies. However, while the scale may reflect the frequency of loans across categories it is not necessarily implicational. Moreover, some of the category designations are ambiguous. In particular, 'free pronouns' can be taken to include both indexical elements (deictic and anaphoric) and indefinites, which however behave quite differently in language contact situations. Thomason and Kaufman (1988: 74ff.) posit that the borrowing of structural categories is gradational and linked to the duration and intensity of social and cultural contacts. Their scale is formulated in general terms of structural properties but allows us to extract the following generalisation in regards word classes: [content words] > conjunctions, adverbial particles > adpositions, personal and demonstrative pronouns, low numerals. While agreeing essentially with earlier observations, the scale does not offer an explanation as to why certain categories should be more resistant to borrowing, and what properties condition ease of borrowing. The link between intensity and duration of contact, and the ordering of categories, is thus associative or even corollary but not causal.

In his discussion of Hungarian loans in Selice Romani spoken in southern Slovakia based on the Loanword Typology Project elicitation list of 1430 lexemes, Elšík (2009) identifies loans as constituting 63% of all nouns on the list, 50% of adverbs, 42% of adjectives, 41% of verbs, and 23% of function words. For the entire database of 41 samples languages considered in the Loanword

Typology Project, the percentage of loans among nouns is 31.2%, while among adjectives and adverbs it is 15.2% and for verbs it is 14.0% (cf. Haspelmath and Tadmor 2009). This way of counting takes into consideration the relative frequency of word classes among lexemes that are deemed to be most salient in everyday speech. The category of 'function words' remains fuzzy and includes both elements that are observed to be highly borrowable such as connectors, particles, and indefinite expressions, and those that are rarely borrowed such as personal pronouns. In an overview assessment Tadmor (2009: 59-63) reports that the database shows twice as many content word borrowings as function words, though there are significant differences among the sample languages: At the far end of the scale, Imbabura Quechua shows a ratio of 14.4 loan content words to loan function words, while Kildin Sami and Selice Romani both show 2.1 and in Berber that ratio is 1.3. In terms of word class frequency, twice as many nouns are borrowed across the corpus than adjectives and adverbs, with the latter closely followed by verbs. Here too there are significant differences: At the far end Takia shows a loan noun to loan verb ratio of 11.8, while in English that ratio is 1.4. The very high borrowers Selice Romani and Berber show loan noun to loan verb ratios of 1.7 and 1.3 respectively, while the very low borrowers Ket and Manange show 3.4 and 3.7, respectively, confirming that heavy borrowing makes the borrowing of verbs much more likely. The statistical comparison between the sample languages renders a borrowability score for each of the words on the elicitation list. Among the items that are least borrowable we find personal pronouns, demonstratives, and place adverbs (there, here), while interrogatives show a range of different scores: Least borrowable are 'where', 'why', and 'which', with scores of 0.997, 0995, and 0.994 respectively; followed by 'how' (0.980), 'what' (0.971), 'when' (0.953) and 'how much' (0.946). This gives, in reverse, the following hierarchy representing the likelihood of interrogative borrowing: 'how much' > 'when' > 'what' > 'how' > 'where', 'why', 'which'.

The split among interrogatives in the Loanword Typology study shows that hierarchies of borrowability might be more meaningful when applied to individual values within a category rather than to the comparison between word classes. Matras (1998) postulates a universal tendency for contrastive connectors to show higher susceptibility to borrowing, followed by disjunctives and then additives ('but' > 'or' > 'and'). Examining a sample of Romani dialects in contact with a variety of different languages, Elšik and Matras (2006: 370-6) show that for some categories values that are 'unmarked' (more frequent, less complex) are more easily borrowed while for other categories the reverse applies. Borrowing is thus found to be linked not to frequency or formal composition but to the conceptual status of category values: lower accessibility and greater semantic-pragmatic complexity (e.g. contrast, free-choice indefinites, non-positive degree, peripheral localisation) correlates with greater susceptibility to borrowing. Matras (2007) presents similar findings in a discussion of a cross-linguistic sample. The high borrowability of category values such as contrast, modality, obligation, and free-choice indicates that susceptibility to borrowing is sensitive to the processing of presuppositions, in particular where a clash is anticipated between the expectations of the speaker and the listener and the speaker needs to exert greater control to manage the interaction (see also Matras 2020: 171ff).

This interactional dimension is particularly relevant when it comes to the borrowing of function words. While earlier hierarchies based on quantitative impressions ranked function words low for borrowability, many studies have flagged the fact that particular kinds of function words are easily borrowed. Salmons (1990) discusses motivations for discourse-pragmatic operations to converge in bilingual speech leading to wholesale borrowing of discourse markers. Van Hout and Muysken (1994) report that Spanish items in Bolivian Quechua include frequently used words in Spanish and few items that are highly inflected, while items that are discourse-related are more susceptible to borrowing than those that belong to the structure of the clause. Stolz and Stolz (1996) examine borrowing in some 30 Mesoamerican languages in contact with Spanish and note the frequent borrowing of function words, in particular discourse markers and connectors. Matras (1998) postulates a functional category of 'utterance modifiers' including connectors, fillers and tags, and phasal adverbs, all of which are particularly prone to borrowing. These observations reinforce doubts as to whether an overall hierarchy of borrowability by word class is useful and indeed whether the distinctive features of sentence-based categories are at all aligned with the factors that drive or hinder borrowing. Matras (2020: 345-6) argues that the high susceptibility to borrowing of categories such as connectors and sequential adverbs, and to a lesser extent phasal adverbs, focus particles, indefinites and modal particles, stems from their role in managing the interaction and processing gaps in presupposition, while at the opposite end of the scale categories such as personal pronouns, demonstratives and place adverbs are less prone to borrowing precisely because the rely on the harmony between speaker in listener in mapping the presuppositional domain. Borrowing is thus regarded as the outcome of lapses in control over the selection and inhibition of structures by language, leading to convergence or 'fusion' of forms.

There thus appear to be at least two separate kinds of motivations for borrowing, affecting different categories in different ways: Lexical borrowing is a process by which means of expressions are enriched. Nouns represent the most obvious way to enrich lexical expression even in the absence of full bilingual competence. Grammatical borrowing by contrast is motivated by the need to ease processing of propositions and the management of communicative interaction, by reducing the burden of having to sustain the selection and inhibition mechanism through which bilinguals make choices (cf. Green 1998; Bialystok et al. 2009). This latter motivation conditions the borrowability of non-lexical items or function words. This helps explain some features of postulated hierarchies of borrowing by word class, such as the prominence of nouns followed by other content words, and of connectors, while pronouns, determiners and pronominal clitics appear low on the hierarchies. The precise interrelations among other categories, for example the position of adpositions or of subordinating conjunctions, is less clear and may be more worthwhile to investigate in regard to the relations between category values than word classes. It is important to note, however, that while some categories appear to be more immune when it comes to the direct transfer of forms or replication of 'matter', some may be more prone to replication of 'pattern' – the mapping of form and meaning (cf. Matras 2020; Matras and Sakel 2007). This is quite obvious for definite and to some extent indefinite articles, which are often found in areal

clusters, that is, prone to convergence, yet are rarely transferred directly from one language to another. It also applies to clitics and modifiers of tense and aspect categories, and distinctions within pronoun paradigms such as inclusive/exclusive, while prepositions, subordinating conjunctions, complements and relativisers can appear as either matter or pattern replicas.

A final note in connection with borrowability concerns the word class distribution by etymology in cases of Mixed Languages (MLs). While opinions differ, by and large MLs are considered to be distinct from cases of borrowing in that they involve the abrupt emergence of a new variety of speech that draws more or less equally on two separate language components which the founder generation had at their disposal as a bilingual community. Bakker (1997, 2003) notes the predictability of structural intertwining in mixed languages, with the etymological split between lexicon and grammar, or nouns and verbs, and function words patterning in different ways, sometimes split between the two sources. According to Matras (2003) the ML 'prototype' is characterised primarily by a split between predication grammar and basic lexicon, while individual word classes can pattern either with the source language of the predication grammar or with that of core lexicon. For example, personal pronouns and demonstratives pattern in Michif with Cree, the donor language of verbs and verb inflection, and in Copper Island Aleut they pattern with Russian, the language of verb inflection; in Gurindji Kriol they pattern with Gurindji, the donor language of nominal inflection, while in Media Lengua they pattern with Spanish, the donor language of the lexicon. But structures involved in coordination, complementation, conditionals, causal relations, negation and relativizing pattern overwhelmingly with the language of finite predication (Matras 2003: 165). This gives us a rough indication that the category of 'function words' is best divided at least between indexical devices (deixis and anaphora) on the one hand and operators that modify the predication, on the other. Where there is a two-way etymological split between components, the first (indexical devices) can pattern with either of the donor languages while the second tend to pattern with the language of finite predication grammar.

3. The integration of borrowed word classes

As discussed, nouns are prominent borrowings, representing a differentiated inventory of labels for concepts, practices, artefacts, products, human agents, and more. As Matras (2020: 188ff.) argues, it is the referentiality of nouns rather than their structural features that motivates the borrowing of nouns. 'Borrowability' should therefore be considered a direct product of communicative, social and pragmatic motivations to borrow rather than a matter of formal constraint. Languages either treat borrowed nouns like native nouns and integrate them into native inflection patterns, or avoid integration and maintain a simplified representation of borrowed nouns. In some cases, nouns are integrated along with their original inflection from the source language, or else a special integration strategy is applied that marks out borrowed nouns as loans. Among the more common replications of source language inflection is the retention of plural marking (English *stimulus-stimuli*), often in addition to the indigenous or inherited plural inflection (Jerusalem Domari *zlām-e* 'men', from

Arabic *zlām* 'men'; cf. Matras 2012), and the adoption of gender distinctions (Malay *putra* 'son' and *putri* 'daughter', from Sanskrit; cf. Tadmor 2007). In both cases the inflection is arguably derivational, thus referencing meaning rather than syntactic role. Swahili famously re-interprets segments of borrowed lexemes as grammatical markers of nominal class, as in *kitabu* 'book' from Arabic, re-analysed to form the Swahili plural *vi-tabu* 'books'. By contrast, in Algerian Arabic the French definite article can be reinterpreted as a plural marker, as in *kadu* 'gift', plural *likadu* 'gifts', while in Spanish the definite article in Arabic loans is reinterpreted as part of the lexical stem, as in **ar-roz* 'rice', **al-calde* 'mayor'. Romani dialects across Europe rely on nominal inflection endings borrowed from Greek during the Early Romani contact period with Byzantine Greek for the integration of both Greek and subsequent loan nouns, for example *doktor-os*, *doktor-is*, *doktor-o* 'doctor'.

Verbs show a variety of morphological integration strategies (cf. Matras 2020: 191ff; Wohlgemuth 2009; Wichmann and Wohlgemuth 2008). They can appear without modification of the original form of the verb ('direct insertion'), as in English *demand*; with morphological modification of the original form of the verb ('indirect insertion'), as in German *analys-ier-en* 'analyse', often drawing on derivational morphology used to create verbs from non-verbs (de-nominal or de-adjectival) or for intensification of an action, as in Hebrew *tilfén* 'to telephone'; in a compound construction where the verb stem or nominalised form is accompanied by an inherited verb ('light verb') as in Turkish *zann-etmek* 'to contemplate' (from Arabic *ẓann* 'think' and Turkish *etmek* 'to do'); or by replicating the original verb along with its original inflection ('paradigm transfer'), as in the replication of Turkish verbs such as *evlendim* 'I got married' in some Romani dialects of the Balkans or the replication of inflected modal and auxiliary verbs from Arabic such as *baqēt* 'I continued' in Jerusalem Domari (Matras 2012). Wichmann and Wohlgemuth (2008) propose that the various strategies represent a hierarchy of the intensity of contact, whereby increasing bilingualism requires less integration effort. Accordingly, light verbs represent the integration strategy that accompanies more casual contacts, followed by indirect insertion, while direct insertion and paradigm transfer are associated with more intense bilingualism. However, examining a database of some 800 examples of loan verbs from over 400 different languages, Wohlgemuth (2009: 286) concludes that the degree of bilingualism and contact intensity cannot fully explain the distribution of integration strategies. The question requires further investigation, but there are indications that the structure of the contact language may play a role. Thus, contemporary young speakers of Kurmanji in Turkey appear to favour a light verb integration pattern with loan verbs from Turkish, as in *beklemiş kir* '(I) waited', while the same generation of Kurmanji speaker in Syria favour indirect integration with Arabic loan verbs, as in *meşîyam* 'I walked'; both groups are exposed to a similar level of bilingualism.

Adjectives tend to be integrated syntactically into the position of the attribute in the recipient language, thus Hebrew *yéled inteligénti* 'an intelligent boy', Yiddish *mešigene mentšn* 'crazy people' (from Hebrew **mešuga*). Like nouns, adjectives often adopt the agreement morphology of the recipient language, thus German *ein cool-er Typ* 'a cool guy', *die cool-en Typen* 'the cool guys'. By contrast, Urdu adopts English adjectives without integrating them into Urdu adjectival inflection, thus *final elān* 'final notice'. In Turkish, which lacks

gender or number agreement, Arabic adjectives are replicated in the default masculine singular form: *ciddi bir plan* 'a serious plan' (Arabic *ǧidd-ī* 'serious-M.SG'). Maltese integrates Italian adjectives into its own gender distinction in the singular, thus masculine *modern*, feminine *moderna*, where the inflectional endings of the two languages happen to coincide, but adopts the Italian inflectional ending in the plural, thus *toroq modern-i* 'modern roads'. Integration into the morphology of the host language usually also applies to comparative and superlative forms of adjectives, thus German *cool-er, am cool-sten* 'cooler, 'coolest'. But cases of wholesale borrowing of non-positive forms of adjectives are also attested: Jerusalem Domari for instance borrows non-positive forms from Arabic even for adjectives for which the language retains inherited positive forms, rendering in effect a system of adjective formation that is consistently suppletive, thus *tilla* 'big', *ákbar* 'bigger', *kištota* 'small', *ázġar* 'smaller' (Matras 2012).

The susceptibility of other word classes to borrowing is often explained by their sentence-peripheral or phrase-peripheral position and absence of inflection. While these may be facilitating factors, we have no evidence that the presence of morphology or morphological boundness actually constitute barriers to borrowing where there is a functional or language-processing motivation for borrowing, and sociolinguistic conditions are such that borrowing is allowed to propagate from isolated innovations in the speech of individuals to become a widespread pattern across a speech community. Thus, connectors and other utterance modifying particles such as focus particles and phasal adverbs ('still', 'already', etc.) are frequent borrowings in forms that are uninflected, unbound and sentence peripheral. But Heath (1978: 100) reports on a number of bound affixes shared by unrelated languages in the Arnhem Land region of northern Australia, including *-ʔŋiriʔ* 'as well as, also', *-bugiʔ* 'only, still'. Jerusalem Domari borrows a direct object resumptive pronoun *iyyāh* from Arabic along with its Arabic inflection, which is used productively in Domari relative clauses, as in *ple illi torim iyyā-hum* 'the money that you have me'. The language also uses the Arabic-derived complementiser *inn-* and subordinating conjunction for cause/reason *li'ann-* along with their Arabic inflections (cf. Matras 2012), whereas in Syrian Kurmanji the same Arabic items are borrowed without Arabic inflection.

The borrowing of indefinites often shows a split between the marker of indefiniteness, which is borrowed, and an inherited ontological marker, thus Polish Romani *ni-so* 'nothing', *vare-kon* 'somebody', where *ni-* is Slavic and *vare-* originally Romanian. Here too borrowablity correlates with category values and the tendency to borrow entire indefinite expressions prevails particular with temporal indefinites wholesale, as in both Syrian Kurmanji and Jerusalem Domari *da'iman* 'always' and *abadan* 'never' from Arabic. As mentioned above, split category values are also attested for interrogatives ('how much' and 'when' being highly borrowable); with prepositions, where facilitating semantic factors include distance complex reference points ('between', 'around'); and with contrastive meaning ('instead of', 'against', 'except for'; see e.g. Elšík and Matras 2006: 287-94).

The borrowing of numerals is sometimes explained as driven by a need to fill gaps in languages that have just rudimentary counting systems that distinguish relative quantities based on a highly presuppositional

conceptualisation of what constitutes 'many' or 'few'. But the replacive borrowing of numerals is not uncommon. Here too, category values play a key role, with borrowing more likely to affect higher and more abstract numerals. This is aligned with frequency and conceptual accessibility but also with sociolinguistic usage patterns, whereby borrowing appears in domains where counting is associated with interaction in the contact language such as negotiation of dates and business transactions (cf. Matras 2020: 217-19). Williams-Van and Hajek (2018) confirm this for the Tetun Dili language of East Timor, where alongside the indigenous set of numerals there is wholesale borrowing of sets from Indonesian and Portuguese. Speakers select sets by interaction domains such as dates and maths (Portuguese), education and prices (Indonesian) and domestic affairs and the family (Tetun Dili). Indonesian and Portuguese numerals, however, can only combine with nouns of the same respective etymology, and their morphosyntax follows the patterns of the respective source language. Jerusalem Domari borrows all numerals about '5' from Arabic and like Tetun Dili, it tends to select Arabic nouns when using Arabic numerals, apparently as a way of avoiding the typological clash among the two languages, as Domari numerals take singular noun while Arabic numerals under '10' take plural nouns, thus *di wars* 'two year-Ø' but *talat snīn* 'three years' (Matras 2012).

Finally, it has been observed that personal pronouns are seldom borrowed (see Tadmor 2009). Exceptions noted in the literature are few and include primarily cases where a pronominal system was enriched in order to accommodate a new inclusive/exclusive distinction, or else cases where the 'pronoun' is in fact a honorific lexical expression rather than a deictic or anaphoric expression (see Matras 2020: 219-25), particularly common in Southeast Asia. The distribution of pronouns and demonstratives in Mixed Languages was referred to above, as was the example of the Domari resumptive pronoun borrowed from Arabic as part of the template of organising relative clauses. From this we can conclude that while there may not be an absolute constraint on the borrowability of pronouns, they tend not to fall among the categories around which bilingual language users are motivated to reduced processing load by eliminating the need to selection and inhibition among languages. This is connected to the fact that deictic and anaphoric elements operate on the basis of a harmonious mapping shared by speaker and listener of the presuppositional domain and so they do not trigger the kind of interaction tension that is associated with processing discontinuity or presuppositional gaps.

4. Word classes in bilingual speech

The study of bilingual speech corpora offers a glimpse into the real-time phenomena that take place at the level of the speaker and may ultimately lead to borrowing at the level of the speech community. Summarizing research on borrowings based on corpora from several language pairs (e.g. French-English, English-Tamil, English-Japanese, English-Igbo), Poplack (2018) reaffirms that the great majority of other-language insertions are single words and that these are overwhelmingly content words like nouns, verbs, adjectives, adverbs, as well as

interjections. In contrast, multi-word stretches are rare and there are no restrictions as far as word classes are concerned, with articles, pronouns, auxiliaries, prepositions and conjunctions being used in addition to the other categories. Based on the analysis of corpora from less-studied languages like Romani, Balkan Slavic, and Ixcatec (Otomanguean), Adamou (2016) confirms that single-word insertions are the most frequent. A further correlation exists between word classes and the overall rates of contact words in bilingual speech corpora: those with over 20% contact words exhibit greater word class variety while those with less than 5% contact words show a limited number of word classes, with nouns being the most prominent word class followed by verbs. More specifically, in two corpora with less than 5% contact words (Ixcatec-Spanish and Balkan Slavic Nashta-Greek) individual tokens from other word classes are used but their rate within their word class is very low. In comparison, in the Romani-Turkish-Greek corpus the analysis reveals that, in addition to nouns (40% in total from either Turkish or Greek, of which 27% from Turkish) and verbs (12% Turkish and 2% Greek) there is a significant number of adverbs (29% from Turkish and 6% from Greek), adjectives (19% Turkish and 10% Greek), and conjunctions (21% Turkish and 12% Greek), but there are no pronouns or determiners from the two current contact languages. In contrast, in the Finnish Romani-Finnish corpus, 72% of all conjunctions used are from Finnish, 27% of particles, 22% of adverbs, and 13% of nouns, 10% of verbs, and 5% of pronouns. In this corpus it appears that Finnish Romani speakers who codeswitch the most do not use many Finnish nouns. Despite inter- and intra-speaker variability, which impact rates of use of individual word classes such as nouns, the statistics indicate that bilinguals conform to the patterns of codeswitching and borrowing that prevail in their bilingual community (cf. Adamou et al. 2016 on bilingual corpora from four Slavic minority languages).

Regarding morphological and syntactic integration, Poplack (2018) observes that single content words typically adopt the morphology of the recipient language. Determiner use and adjective placement also align with the grammar of the recipient language. This is not the case in multi-word stretches, where words are governed by the grammar of the donor language. Other corpus studies confirm that the choice of the determiner depends on the choice of the 'matrix language', defined by Myers-Scotton (2002) as the language that provides clause word order (Morpheme Order Principle) and morphemes such as case and verb agreement markers (System Morpheme Principle). Blokzijl, Deuchar, and Parafita Couto (2017) compare a bilingual corpus of Spanish-English from Miami, USA and a bilingual corpus of Spanish-English creole from Nicaragua and report a tendency for Miami bilinguals to use Spanish determiners in mixed determiner phrases while Nicaraguan bilinguals use English creole determiners consistently. Adamou and Granqvist (2015), however, note typologically rare patterns of mixing in two Romani corpora from Greece and Finland: Romani-Turkish speakers from Greece never use Romani verb morphology with the Turkish verbs even though Turkish verbs appear as single-word insertions into an otherwise Romani-dominant speech that includes nouns and pronouns with Romani case marking, Romani determiners, and Romani word order. A similar tendency is noted for Finnish verbs used by Finnish Romani bilingual speakers.

From a cross-language perspective careful attention to methodological issues is therefore called for: Some languages have definite and indefinite articles while others express definiteness through bare nouns; some languages rely more and others less on conjunctions, and some have more and others fewer adjectives; in some languages pronouns can be omitted in some environments or contexts while in others they are obligatory; languages that are morphologically isolating, agglutinative, or fusional can offer different opportunities and constraints on the borrowing of morphemes. At the same time, semantic-pragmatic and cognitive motivations may override constraints on the borrowability of bound morphemes, as in the case of the borrowing between Sakha (Yakut) and Lamunkhin Even alluded to above (Pakendorf 2019). Differences in word order can also play a role (affecting for example the borrowing of pre- and postpositions). Finally, the fuzziness of some word classes, alluded to above in relation to 'pronouns' and 'adverbs', makes cross-language comparability difficult. Indeed, Gomez Rendon (2008) argues based on a comparative study of three bilingual corpora from Latin America (Spanish with Quichua, Guaraní, and Otomí), that what ultimately determines the distribution of word classes in different language pairs are the particular characteristics of the bilingual communicative setting rather than the typology of parts of speech and morphology of the recipient language or source language.

To account for the differences in the word classes noted in bilingual speech corpora we now turn to discuss models of speech production. The classic model of monolingual speech production elaborated by Levelt (1989) involves three subsystems: A pre-linguistic conceptual system (Conceptualizer), a linguistic system at the lemma level (Formulator), and an output system at the word form level (Articulator). According to the bilingual version of this model proposed by de Bot (1993), the Conceptualizer first decides on the language that needs to be selected, based on the context. The intended meaning and the context then activate subsets of words belonging to one of the two languages. The Formulator then gains access to the lemma and lexeme levels as in the monolingual speech production model, before moving to the Articulator, which according to de Bot is not language specific; this can explain the variation in the phonetic-phonological integration of words in codeswitching observed in many bilingual speech studies.

Building on Levelt's model, Myers-Scotton and Jake (2009) propose the 4-M model to explain differences in the ways that various morpheme types are used in bilingual speech. The model distinguishes two types of morphemes: conceptually-activated morphemes, which are activated early, at the level of the Conceptualizer, at the same time as content words ('early system morphemes'); and structurally-assigned morphemes, which are activated later in this process, at the level of the Formulator ('late system morphemes'). Early system morphemes include determiners, derivational prepositions, particles in phrasal verbs, derivational and plural markers (the latter being the only kind of inflectional morphology to behave differently in codeswitching and borrowing), some tense and aspect markers, as well as subordinating and coordinating conjunctions. These morphemes can come either from the language of the content words or from the language that provides the grammar. Late system morphemes, in contrast, such as agreement and several case markers, are rarer

in codeswitching and borrowing as they are not activated based on content but are accessed at a later stage in speech production.

Alternative models move beyond the word level and therefore, in some sense and to some extent, beyond word classes (e.g. the 'slot and insert' model in Dell et al. 1997; cognitive grammar in Langacker 2008; naive discrimination theory in Baayen et al. 2013). According to usage-based approaches, in particular, speakers store larger units that can include a word together with its inflectional morphology or even entire lexico-grammatical constructions. Researchers investigating bilingual speech therefore note that some units, such as compound nouns, adjective-noun and verb-object constructions (Backus 1996, 2003), nouns together with plural markers (Hakimov 2016), or verbs together with verb morphology (Adamou and Shen 2019) can also be stored as a whole in the mental lexicon and be accessible as such in bilingual speech production and comprehension, depending on their frequency.

Finally, researchers note that specific elements in a word class or sub-classes are affected differently in bilingual speech production. For example, Backus (1996) observes that proper nouns differ from other nouns in that they have a more specific meaning, which makes them more susceptible to borrowing. Matras (1998) also notes that grammatical markers that are associated with interaction management and so with more intense language processing may be more prone to borrowing (e.g. the adversative marker 'but'). The reason for this, it is hypothesised, lies in the interaction-level tension surrounding the act of contradicting a shared presupposition, and the resulting increase in processing load, which can interfere with the language selection mechanism.

5. Conclusion

Studies in contact linguistics often make reference to word classes, both in respect of the likelihood that different word classes may be affected by borrowing or constitute points of language switching in conversation, and in respect of the characteristics of structural integration into a recipient language of items belonging to different word classes. The results of various observations are difficult to calibrate into overall generalisations, however. This is partly due to the great variety observed in different corpora, in particular those of bilingual speech. There are also obstacles that are of an epistemological nature. Thus, statistics may reflect different realities such as the overall quantity of borrowed or switched items of particular categories in a corpus, their frequency as a proportion of the overall number of borrowings or switches, their frequency in proportion to the overall count of items belonging to the same category, or the implicational arrangement in regard to the sequence of borrowing in the history of contact between two languages, to name but some. Then there is the potential ambiguity of some category labels and, in regard to attempts to derive from the data an explanatory account, the reasoning that is linked to the borrowability of certain categories when these are defined in terms of their sentence level and morpho-syntactic characteristics as word classes. Indeed, one of the outcome observations is that language contact serves as a prompt to re-think the status of categories and of the relevance of the very notion of word classes. Firstly, we find

that the hierarchical nature of borrowing manifests itself particularly around difference susceptibilities of single values within a category, alongside or perhaps even more so than among word class categories. Furthermore, borrowing and switching both seem to illuminate links between categories that are anchored in their functional contribution to the management of interaction and the processing of discourse and thus do not strictly pertain to their status as word classes in the conventional sense. The general observation that content words are frequent borrowings, for example, has to do with their role to enriching content: nouns appear at the top of the list thanks to their contribution to naming objects and concepts, which are key elements of the process of cross-cultural exchange that is inherently involved when languages are in contact. Other elements, such as indefinites, discourse markers and connectors, interjections and greetings, phasal adverbs and fillers, appear high on the borrowing and switching scale due to the processing load that they impose on the speaker as part of the apparatus of 'managing and directing' the interaction and the link to broken presuppositional chains (for a discussion see Matras 1998, 2020). This suggests that psycholinguistic dimensions rather than word class affinity in the traditional sense are the motivating factors behind the behaviour of different categories in contact situations. An agenda for future research should therefore in our opinion see diachronic and corpus based work supplemented by experimental research to investigate the interplay between speaker control over the selection and inhibition mechanism and longer term re-drawing of 'language boundaries' within speakers' repertoire of linguistic resources.

References

Adamou, Evangelia (2016). *A Corpus-Driven Approach to Language Contact: Endangered Languages in a Comparative Perspective*. Berlin: Mouton de Gruyter.

Adamou, Evangelia, Walter Breu, Lenka Scholze, and Rachel X. Shen (2016). 'Borrowing and contact intensity: A corpus-driven approach from four Slavic minority languages'. *Journal of Language Contact* 9(3): 515–44. https://doi.org/10.1163/19552629-00903004

Adamou, Evangelia and Kimmo Granqvist (2015). 'Unevenly mixed Romani languages'. *International Journal of Bilingualism* 19: 525–47. https://doi.org/10.1177/1367006914524645

Adamou, Evangelia and Rachel X. Shen (2019). 'There Are no Language Switching Costs when Codeswitching is Frequent'. *International Journal of Bilingualism* 23: 53–70. https://doi.org/10.1177/1367006917709094

Baayen, Harald R., Peter Hendrix, and Michael Ramscar (2013). 'Sidestepping the Combinatorial Explosion: An Explanation of N-Gram Frequency Effects Based on Naive Discriminative Learning'. *Language and Speech* 56(Pt 3): 329–47. https://dx.doi.org/10.1177/0023830913484896

Backus, Ad (1996). *Two in One. Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: Tilburg University Press.

Backus, Ad (2003). 'Units in Code switching: Evidence for Multimorphemic Elements in the Lexicon'. *Linguistics* 41: 83–132.

Bakker, Peter (1997). *A Language of our Own. The Genesis of Michif – the Mixed Cree-French Language of the Canadian Métis.* New York: Oxford University Press.

Bakker, Peter (2003). 'Mixed Languages as Autonomous Systems', in Yaron Matras and Peter Bakker (eds), *The Mixed Language Debate. Theoretical and Empirical Advances.* Berlin: Mouton de Gruyter, 107–50.

Bialystok, Ellen, Fergus I. M. Craik, David W. Green, and Tamar H. Gollan (2009). 'Bilingual Minds'. *Psychological Science in the Public Interest* 10: 89–129.

Blokzijl, Jeffrey, Margaret Deuchar, and M. Carmen Parafita Couto (2017). 'Determiner Asymmetry in Mixed Nominal Constructions: The Role of Grammatical Factors in Data from Miami and Nicaragua'. *Languages* 2: 20. https://doi.org/10.3390/languages2040020

de Bot, Kees (1992). 'A Bilingual Production Model: Levelt's Speaking Model Adapted'. *Applied Linguistics* 13(1): 1–24.

Dell, Gary S., Lisa K. Burger, and William R. Svec (1997). 'Language Production and Serial Order: A Functional Analysis and a Model'. *Psychological Review* 104: 123–47. http://dx.doi.org/10.1037/0033-295X.104.1.123

Elšík, Viktor (2009). 'Loanwords in Selice Romani, an Indo-Aryan language of Slovakia', in Martin Haspelmath and Uri Tadmor (eds), *Loanwords in the World's Languages. A Comparative Handbook.* Berlin: Mouton de Gruyter. 260–303.

Elšík, Viktor and Yaron Matras (2006). *Markedness and Language Change: The Romani Sample.* Berlin: Mouton de Gruyter.

Gomez Rendon, Jose A. (2008). *Typological and Social Constraints on Language Contact: Amerindian Languages in Contact with Spanish*. Utrecht: LOT.

Green, David W. (1998). 'Mental Control of The Bilingual Lexico-Semantic System'. *Bilingualism: Language and Cognition* 1: 67–81. https://doi.org/10.1017/s1366728998000133

Hakimov, Nikolay (2016). 'Explaining Variation in Plural Marking of Single German Noun Insertions in Russian Sentences', in Harald Beherens and Stefan Pfänder (eds), *Experience Counts: Frequency Effects in Language*. Berlin: Mouton de Gruyter, 21–260.

Haspelmath, Martin and Uri Tadmor (2009) eds. *Loanwords in the World's Languages. A Comparative Handbook.* Berlin: Mouton de Gruyter.

Haugen, Einar (1950). 'The Analysis of Linguistic Borrowing'. *Language* 26: 210–231.

Heath, Jeffrey (1978). *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute for Aboriginal Studies.

Langacker, Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Levelt, Willem J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Matras, Yaron (1998). 'Utterance modifiers and universals of grammatical borrowing'. *Linguistics* 36: 281–331.

Matras, Yaron (2003). 'Mixed Languages: Re-Examining the Structural Prototype', in Yaron Matras and Peter Bakker (eds), *The Mixed Language Debate. Theoretical and Empirical Advances*. Berlin: Mouton de Gruyter.151–75.

Matras, Yaron (2007). 'The Borrowability of Grammatical Categories', in Yaron Matras and Jeanette Sakel (eds), *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter, 31–74.

Matras, Yaron (2012). *A Grammar of Domari*. Berlin: Mouton de Gruyter.

Matras, Yaron (2020). *Language Contact* [second edition]. Cambridge: Cambridge University Press.

Matras, Yaron and Jeanette Sakel (2007). 'Investigating the Mechanisms of Pattern-Replication in Language Convergence'. *Studies in Language* 31: 829–65.

Moravcsik, Edith (1978). 'Universals of Language Contact', in Joseph H. Greenberg (ed), *Universals of Human Language*. Stanford: Stanford University Press, 94–122.

Myers-Scotton, Carol (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.

Myers-Scotton, Carol and Janice L. Jake (2017). 'Revisiting the 4-M Model: Codeswitching and Morpheme Election at the Abstract Level'. I*nternational Journal of Bilingualism* 21: 340–66. https://doi.org/10.1177/1367006915626588

Pakendorf, Brigitte (2019). 'Direct Copying of Inflectional Paradigms: Evidence from Lamunkhin Even'. *Language* 95: e364–80. https://doi.org/10.1353/lan.2019.0063

Poplack, Shana (2018). *Borrowing*. Oxford: Oxford University Press.

Salmons, Joe (1990). 'Bilingual Discourse Marking: Code Switching, Borrowing, and Convergence in Some German-American Dialects'. *Linguistics* 28: 453–80.

Stolz, Christel and Thomas Stolz (1996). Funktionswortentlehnung in Mesoamerika Spanisch-amerindischer Sprachkontakt (Hispanoindiana II). *STUF - Language Typology and Universals* 49: 86–123.

Tadmor, Uri (2007). 'Grammatical Borrowing in Indonesian', in Yaron Matras and Jeanette Sakel (eds), *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: de Gruyter, 301–28.

Tadmor, Uri (2009). 'Loanwords in the World's Languages: Findings and Results', in Martin Haspelmath and Uri Tadmor (eds), *Loanwords in The World's Languages: A Comparative Handbook*. Berlin: Mouton de Gruyter, 55-75.

Thomason, Sarah G. and Terrence Kaufman (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

van Hout, Roeland and Pieter Muysken (1994). 'Modeling Lexical Borrowability'. *Language Variation and Change* 6: 39–62.

Weinreich, Uriel (1953). *Languages in Contact*. The Hague: Mouton.

Whitney, William D. (1881). 'On Mixture in Language'. *Transactions of the American Philological Association* (1869-1896) 12: 1–26.

Wichmann, Soeren and Jan Wohlgemuth (2008). 'Loan Verbs in A Typological Perspective', in Thomas Stolz, Dik Bakker, and Rosa Salas Palomo (eds), *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romanisation Processes*. Berlin: Mouton de Gruyter, 89–121.

Williams-Van Klinken, Catharina and John Hajek (2018). 'Mixing Numeral Systems in Timor-Leste', in Antoinette Schapper (ed), *Contact and Substrate in the Languages of Wallacea* Part 2. NUSA 64: 65–94.

Winford, Donald (2010). 'Contact and Borrowing', in Raymond Hickey (ed), *The Handbook of Language Contact*. Oxford: Wiley Blackwell, 170-87.

Wohlgemuth, Jan (2009). *A Typology of Verbal Borrowings*. Berlin: Mouton de Gruyter.