Borrowing

Yaron Matras, Evangelia Adamou

Borrowing

Yaron Matras and Evangelia Adamou

1. INTRODUCTION/DEFINITIONS

'Borrowing' is a metaphor that denotes the use of a structure (i.e., phone, phoneme, morpheme, semantic value, or form-function alignment) within a particular linguistic system although it is normally associated with another linguistic system. The term 'borrowing' is therefore based on the underlying formal or structuralist assumption that languages are self-contained systems, and that the use of linguistic structures can and should be described in terms of their system affiliation. That assumption has, however, become ever more controversial in recent years, as scholars of contact linguistics refer to users' 'repertoires' (Matras, 2009; Blommaert and Backus, 2013) and stress that bilinguals may interact in a 'bilingual mode' where they may choose to combine elements from both languages (Grosjean, 1989). Support for this view comes from neurolinguistic and psycholinguistic studies showing that bilingual speakers do not 'switch off' individual linguistic systems (e.g., Thierry and Wu, 2007; Kroll et al., 2008; Loebell and Bock, 2003). Moreover, cross-discipline perspectives in linguistics, ethnography, and education have taken an interest in what is sometimes referred to as 'metrolingual' or 'translanguaging practices,' understood as users' ways of availing themselves of the full expressive potential of their entire repertoire of linguistic features in communicative interaction (Li Wei, 2018). As a result, there is a need to distinguish between multilinguals availing themselves of structures that form part of their overall repertoire of linguistic features, and an etymological analysis of individual items

(words and structures). Depending on the approach, borrowing can therefore have either a synchronic or diachronic reading and be the object of investigation in the field of bilingualism and language contact, respectively.

Other terms that are used in connection with the crossing of language boundaries are 'loans,' 'transfers,' 'copying,' and 'replication.' Matras and Sakel (2007b), in particular, refer to two kinds of replication of a linguistic structure: replication of linguistic 'matter' (concrete, identifiable sound-shapes of words and morphs) and replication of mode of organising units of speech, that is, of linguistic 'patterns.' The languages involved in this process are commonly referred to as 'model' and 'replica' (see Heine and Kuteva, 2005).

2. HISTORICAL OVERVIEW

In the nineteenth century, Whitney (1881) responded to contemporary discussions on the possible existence of 'mixed languages' and, in particular, to the question of whether mixture is possible in grammar. He argued that there is nothing in the nature of language that could stand in the way of appropriation of word forms from another language, and introduced the term 'borrowing' to refer to such cases (Whitney, 1881: 10). Whitney emphasised the unequal role of the two languages in contact as, for example, in the case of English and Norman French, inflection was not borrowed together with verbs and nouns; in other words, grammar was exempted from mixture, and this also applied to pronouns, articles, prepositions, conjunctions, and numerals. Whitney further addressed the challenge to bring empirical proof that shows otherwise and to explain the conditions under which such processes occur (Whitney, 1881: 15). It was noted, however, that grammatical material may be borrowed along with words, as in the borrowing of productive word formation affixes with French word

forms, something that allows for inflection to take the same route, as in the plural ending in the word *phenomena* (Whitney, 1881: 17). These were dubbed 'secondary processes.' Moreover, Whitney pointed out the possibility for analogous borrowing as in English *one* from French *on*, and the borrowing of *second*. He further noted the need for a scale of the comparative ease and difficulty of 'immediate borrowing' (Whitney, 1881: 19). He postulated such a scale in the following way: noun > adjective > verb (where > indicates that a category is more easily borrowed). Whitney (1881) also discussed the typological conditions that could facilitate borrowing, such as conversion possibilities among English word forms that facilitate the borrowing of verbs and adjectives. By contrast, Persian does not borrow Arabic verbs, but instead uses nouns to form quasi-verbs. Finally, Whitney (1881) addressed the replicability of morpho-syntactic patterns such as the order of heads and attributes and suggested that more formal or structural material is more resistant to borrowing.

The first half of the twentieth century is another key moment in the history of contact linguistics, with several authors focusing on borrowing. Haugen (1950), for example, postulated that the analysis of borrowing must begin with an analysis of the behaviour of bilingual speakers. He pointed out the awkwardness of the metaphor 'borrowing' even though its use as a technical term has remained largely unambiguous in linguistics. Borrowing is essentially the reproduction by speakers of new linguistic patterns in the context of a language other than the one in which those patterns were acquired. Every borrowing must have at some point appeared as an innovation, whether or not the speaker is conscious of that. Borrowing is thus a process of 'innovative reproduction' (Haugen, 1950: 212). Haugen distinguishes 'importation,' understood as the complete replication of material from another language, from 'substitution,' understood as replication that involves alteration of aspects of the structure of the borrowed item. He also reviews various competing and closely-related terms such as

'loanwords', which he distinguishes from 'hybrids' (where only a part of the word is borrowed), 'loan translations' or 'calques' (where form-function mapping is replicated), and 'semantic loans' (where word semantics are reproduced). In particular, Haugen introduces three categories: 'loanwords' (importation but no morphemic substitution), 'loanblends' (importation and partial morphemic substitution of certain features), and 'loanshifts' (complete morphemic substitution without importation, shift in semantic meaning of an existing word based on similarity with external words), including 'loan translations' (words that come into being through contact but are not directly imported from or even modelled on a specific item from the contact language). As a contemporary example of the latter (unknown to Haugen), we can identify today the German expressions *Handy* 'a mobile phone, cell-phone' and *Beamer* 'a projector linked to a computer', which replicate English-derived word-forms, but with novel meanings. Haugen also reviews Whitney's scale of borrowing, acknowledging that all features can be borrowed, but that there is a 'scale of adoptability' that is correlated with structural organisation, but does not elaborate on that point. He does, however, introduce the need for cross-linguistic research, hinting that there may be differences in borrowing among languages. Moreover, he discusses the effects of borrowing on a given linguistic system, noting that it leads to instability. Finally, he discusses phonological, phonetic, and morphological adaptation. In sum, Haugen (1950) largely sets the research agenda on borrowing by addressing key issues such as the manner of structural integration, the structural impact of borrowing, the existence of a universal scale of borrowability of categories, and cross-linguistic differences in borrowing patterns.

At roughly the same time, Weinreich (1953) highlighted the fact that individuals are the locus of contact, directing attention to the fact that languages are said to be in contact if they are used alternately by the same person. Weinreich focuses on 'interference' among languages,

where elements that do not belong to the main language of a speech event are considered to be borrowed or transferred. For Weinreich, interference is the transfer within an act of speech. Much of his discussion is devoted to an overview of types of transfer in phonetics, grammar, and lexicon. He notes that transfer of borrowed morphemes is rare and that the fuller the (structural) integration of a morpheme, the less likely transfer is to occur. According to Weinreich (1953: 35), nouns show high borrowability because of their semantic function as designations of new things. Borrowing of lexicon is also conditioned by the frequency of words, where there is a need for affective enrichment or euphemism, a need for differentiation, and a general need for renewal. Weinreich also distinguishes 'replica' from 'transfer' (1953: 37ff) which corresponds to interference in the domain of grammatical relations such as word order, or agreement, or functional extension.

In the second part of the twentieth century, Moravcsik (1978) proposed a link between structural autonomy, semantic transparency, and borrowability. This results in the following implicational hierarchies: nouns > verbs, derivational > inflectional, unbound > bound. Heath (1984) distinguishes convergence from transfer of forms. He also raises questions regarding adaptation, resulting from conflicting phonological patterns. He introduces the concept of 'routines' to denote the habitual adaptation pattern of borrowed forms. Together these contributions set the agenda for structural types of borrowing, structural patterns of integration, motivations and societal conditions for borrowing.

3. CRITICAL ISSUES AND TOPICS

Distinguishing borrowing from codeswitching

A major topic of discussion in the literature is the distinction between borrowing and codeswitching insertions. There are two main strands of analysis: some researchers consider that borrowings are distinct from codeswitching, others that the distinction is not clear-cut but is best understood as a continuum.

The proposal that codeswitching and borrowing are two different phenomena is expressed by Poplack, Sankoff, and Miller (1988), Poplack and Dion (2012) and in a different theoretical frame by MacSwan (2016) (also see Chapters 3 and 5). Whereas Poplack, Sankoff, and Miller (1988) assess various criteria to define 'borrowing,' Poplack and Dion (2012) suggest that grammatical integration is the best diagnostic criterion and that it correlates with the degree of composition. In this view, single-word tokens tend to be treated like borrowings as they are integrated into the grammatical frame of the recipient language, unlike multi-word codeswitching insertions. Building on the Minimalist Program, MacSwan (2016) proposes a lexicalist framework. In this approach, bilinguals have two lexicons that include grammatical aspects, and two phonetic forms that follow two phonological systems. In this perspective, phonological integration is the best diagnostic feature of borrowings, covering lexical borrowings that are widespread in the community as well as 'nonce borrowings,' that is, borrowings used only once by a given speaker.

In contrast, researchers such as Myers-Scotton (1993) and Matras (2009) consider the existence of a borrowing-codeswitching continuum. For Myers-Scotton (1993), single words are codeswitching insertions with the potential of becoming borrowings as they become

conventionalized, that is, shared by various members of a speech community. There is, according to that view, no structural distinction between borrowing and codeswitching, merely a social distinction between usage conventions. Matras (2009: 111) considers a variety of diagnostic features to determine an item as a borrowing or as a codeswitching: degree of speaker bilingualism (monolingual vs. bilingual), degree of item composition (utterance vs. single lexeme) and of functionality (stylistic vs. default use), unique character of the referent (lexical vs. para-lexical), operationality (core vocabulary vs. grammatical operations), regularity of the process (single vs. regular occurrence), and structural integration (non-integrated vs. integrated).

Integration

For many researchers, the defining characteristic of a borrowing is its integration into the phonology, morphology, syntax of the recipient language and modulation of their semantic scope with respect to the recipient language's equivalent terms. It is generally considered that if an item is integrated by means of phonology, morphology, and syntax, then it can be considered a borrowing.

Standard phonological models assume that the integration of borrowings is a phonological process that takes place in production (Hyman, 1970). More recently, however, it has been proposed that integration also takes place in perception, at a pre-phonological level (Peperkamp and Dupoux, 2003; Yip, 2006). Peperkamp and Dupoux, in particular, suggest that phonetics most likely drive the integration process based on the first language's 'phonetic decoder.' It is interesting to note that borrowed items may show various degrees of phonological integration. For example, Adamou and Arvaniti (2014) report that Romani-Turkish bilinguals approximate some Turkish vowels that do not belong to the native Romani

inventory such as [y], [ɯ] and [œ] together with the Turkish words. On the other hand, Arabic loans in Turkish are often exempted form the rules of (grammatical) vowel harmony, thus the plural formation of *harf* 'letter' is *harf-ler*, not **harf-lar* (Al-Hashmi, 2016: 32). Which phonetic/phonological phenomena will be adapted to the phonetics and the phonological rules of the recipient language seems to depend on a combination of perceptual and articulatory ease. Moreover, as Matras (2009) notes, other motivations may play a role, as, for example, bilinguals familiar with the pronunciation of a word in the donor language may choose to 'authenticate' it by approximating its original phonology.

Structural factors may sometimes drive the choice of the integration strategy, though structural equivalence between two languages is not a pre-condition for borrowing. For example, Domari speakers borrow the full Arabic word-form for all comparative/superlative forms of adjectives, resulting in complete borrowing-based suppletion of the inventory of adjectives: *tilla* 'big' native Domari with *ákbar* 'bigger' from Arabic (Matras, 2009). The motivation behind this choice appears to be that the Arabic morpho-phonological template that forms the comparative/superlative, namely áCCaC, is not easily integrated into the morphological structure of Domari, yet there is at the same time a strong cognitive motivation for the comparative/superlative procedure to undergo fusion with that of the contact language Arabic.

In a cross-linguistic study of verb borrowing integration with examples from over 400 languages Wohlgemuth (2009) identified four major strategies: (a) verb integration through light verbs, (b) indirect insertion, that is, verb integration through affixes, (c) direct insertion without any morphological or syntactic accommodation, and (d) paradigm transfer, for verbs borrowed together with their original verb morphology. Light verbs are commonly used

cross-linguistically. For example, Spanish verbs are integrated into the Otomanguean language Ixcatec with the Ixcatec verbs *tse* 'do' or *tsu* 'want.' The light verbs receive the person and tense-mood-aspect markers and are followed by the Spanish verb stem, e.g., *tse pregúnta* 'ask' (from Spanish *pregunta-r*) *tse ʔadbertí* 'warn' (from Spanish *adverti-r*) (Adamou, 2016: 90). In comparison, an example of direct insertion of Spanish verbs comes from Imbabura Quechua where Quechua verbal inflection is added directly to the Spanish stems, e.g., *balura-ni* 'I value' (from Spanish *valora-r*) (Gómez-Rendón, 2007). Indirect insertion of loan verbs may take place through affixes. This is common in Romani where inflectional or derivational affixes from Greek, such as the Greek aorist *-is-/-as-/-os-* and the Greek present tense *-iz-/-az-/-oz-* and *-in-/-an-/-on-*, have lost their grammatical value and serve as 'loan verb markers' (Elšík and Matras, 2006: 324–333). These affixes were originally used with Greek verbs, as in *ir-in-av* or *ir-iz-av* 'I return' (from Greek *jir-íz-o*), and for past tense *ir-is-ájlom* 'I returned' (from Greek *jír-is-a*). They were then extended to verb borrowings from other contact languages, as in *misl-in-av*, *misl-iz-av* 'I think' (from Slavic *misl-*). Finally, paradigm transfer is noted among various Romani dialects spoken in Russia, Finland, and the Balkans. For example, in Romani spoken in Greek Thrace, Turkish verbs consistently take Turkish verb morphology, that is, various tense-mood-aspect markers, but also all person markers, as well as the negation and causative affixes (Adamou, 2016).

To account for preferred verb-integration strategies across the languages of the world, Wohlgemuth (2009) assesses the hypothesis that these could correlate to the degree of contact and bilingualism. Although paradigm transfer is indeed a strategy that requires a high degree of bilingualism, this hypothesis is ultimately ruled out. Matras (2009) proposes instead that the choice of inflectional system in which to anchor the predication is crucial as it symbolises the bilingual speaker's context-bound choice of 'language.' In that sense, some speakers

integrate verbs merely as lexical labels while others use them, to various degrees, as predicate-initiating devices. More specifically, paradigm transfer is possible when there is full acceptance of a bilingual group identity that licenses speakers to initiate the predication in either language. An alternative account is offered by Myers-Scotton and Jake (2014) who hypothesize that verb integration into the recipient language's morphology should be preferred as being less costly, since the bilingual would only need to control the semantic-pragmatic features and not to check the congruence between the two languages. However, psycholinguistic evidence from Turkish-Romani bilinguals who make use of the paradigm transfer strategy shows that processing morphologically non-integrated verbs in comprehension does not incur cost as long as this has become a habitual verb-integration strategy (Adamou and Shen, 2019).

Borrowability

'Borrowability' is a term that refers to the susceptibility of categories to borrowing. It is often expressed through borrowability scales and implicational hierarchies. Borrowability scales are primarily associated with the frequency of borrowed items by category in a cross-linguistic perspective. Implicational hierarchies, by contrast, suggest that borrowing follows a specific development, where some stages precede other stages in a non-trivial manner.

Thomason and Kaufman (1988) were the first to propose a link between structural categories of borrowing and the duration and intensity of cultural contact ranging from 'casual contact' and lexical borrowing, to 'very strong cultural pressure' and heavy structural borrowing (Thomason and Kaufman, 1988: 74–75). While very influential, the proposal suffers from a number of shortcomings: the categories on the scale remain somewhat vague (e.g., 'function words') and no criteria are provided to quantify the degrees of intensity of contact (e.g.,

'slightly more intense contact'). More importantly, no explanatory account is offered for the link between category susceptibility to borrowing and intensity of cultural contact.

Campbell (1993) offered an evaluation of various proposed generalisations in the contact linguistics literature. While these did not address general patterns of borrowing, they have contributed to the debate on the status of counter examples. Campbell criticises the notion of structural compatibility as a pre-requisite for borrowing. He also criticises the notion of gaps as a motivating factor for borrowing, that is, the idea that the borrowing of a morpheme necessarily leads to the replacement of the corresponding inherited morpheme rather than overall system enrichment. He further criticises the claim that free standing morphemes are more borrowable than bound morphemes, citing individual examples of borrowed bound morphemes (including the mixed language Ma'a, based on Thomason and Kaufman, 1988). Campbell (1993) also critiques Moravcsik's stated universals (Moravcsik, 1978: 101ff) noting that the primacy of lexical borrowing as a pre-requisite for non-lexical borrowing may be true empirically, but is not theoretically significant since there is no link between the two. With support from some empirical evidence, he also rejects the generalization that verbs cannot be borrowed, and argues that despite being common, the borrowing of derivational morphology is not an absolute a pre-requisite for the borrowing of inflectional morphology, noting that borrowing of plural markers such as Spanish -*s* is found in Quechua without any borrowing of derivational markers. Instead Campbell slightly reformulates this implicational hierarchy, saying that it is unlikely for semantically weak affixes to be borrowed (Campbell, 1993: 103). Finally, he offers counter-examples to the notion that grammatical items can only be borrowed along with their linear rule as, for instance, Arabic prepositions are borrowed as postpositions in Turkish (e.g., *kadar*).

Evidence from two cross-linguistic samples have significantly enriched this discussion. The first sample consists of data from 27 languages collected through responses to a typological questionnaire that was filled in by language specialists (Matras and Sakel, 2007a). The second sample comes from the Romani Morphosyntax Database bringing together responses to a questionnaire from over 75 Romani dialects in contact with 25 different languages (Elšík and Matras, 2006). The analysis of these samples confirms a connection between susceptibility to borrowing, and the truth- or presupposition value assigned by a category to propositional content. A good example is the borrowability hierarchy of connectives 'but > or > and,' as well as the hierarchy 'modality > aspect > tense,' 'indefinites > other pronouns,' 'prosody > segmental phonology,' as well as the tendency for more complex or discontinuous local relations such as 'against' and 'except' to show higher borrowability than basic local relations such as 'on' or 'at.'

Motivations for borrowing

Motivations for borrowing can be defined as the factors that drive the speaker to use linguistic matter from one language in interaction settings that are generally associated with another language.

Traditionally, motivations for borrowing were conceived of as falling under two types: filling a 'gap' (typically including cultural loans that are associated with novel activities and artefacts) and reflecting 'prestige' (covering the use of items from the language of a more powerful community to achieve a wide range of social outcomes). However, recent research focuses instead on a wider variety of factors that can be summarized as follows: the need to provide labels for unique referents; association with specific activity domains; the cognitive pressure resulting from the manipulation of the presupposition domain or management of the

12

interaction roles, and, in particular, the clash between speaker intentions and hearer expectations (Matras, 2009).

Van Hout and Muysken (1994) investigated the use of Spanish borrowings in Bolivian Quechua and found that Spanish items that were discourse-related were more susceptible to borrowing than items related to the structure of the clause; that speakers tended to borrow fewer items that are highly inflected; and that frequently used words in Spanish are more likely to be used in Quechua. Stolz and Stolz (1996) examined borrowing in roughly 30 Mesoamerican languages in contact with Spanish and noted the systematic borrowing of function words. They concluded that structural criteria or gaps are not relevant explanatory factors, but that the pragmatics of communication offer the best explanatory account. Salmons (1990) further elaborated on the pragmatic motivations of borrowing, later expanded by Matras (1998, 2009) to relate borrowings to the communicative interaction setting and to the effects that the choice of linguistic items from a given language is expected to have on the interlocutor. From this perspective, gaps are not seen as deficiencies of a given language's expressive means. Rather, borrowings offer bilinguals the possibility to make use of their full inventory of linguistic resources by lifting the restrictions in language selection and using linguistic items in a wider variety of interactional settings. Matras (1998) more specifically accounts for this preference by referring to the role of cognitive factors, that is, to the effort that is needed to keep the subcomponents of a bilingual's language repertoire apart. Evidence from cross-linguistic sampling of grammatical borrowing indicates that grammatical markers associated with complex language processing may be more prone to borrowing (as in the case of the adversative marker 'but'). The explanation offered in Matras (1998) is that the interaction-level tension surrounding the act of contradicting a shared presupposition may interfere with the language selection mechanism or 'executive control.'

13

Another influential account that takes into consideration cognitive factors comes from the 4-M model building on Levelt's speech processing model (see Myers-Scotton and Jake, 2017 and Chapter 4). According to the 4-M model, the conceptually-activated morphemes (the 'early system morphemes') are more susceptible to borrowing as they are salient in the mental lexicon along with their content morpheme heads; this is the case of determiners, derivational prepositions, and particles in phrasal verbs, as well as certain affixes (derivational and plural markers, some tense and aspect markers, some subordinating and coordinating conjunctions). In contrast, the structurally-assigned morphemes (the 'late system morphemes') are rarely borrowed because they carry little content and are accessed late in speech production; this is the case of 'bridges' that join together two NPs and complementizers that join together two clauses, and 'outsiders' that include the agreement markers and several case markers.

Borrowing and language genealogy

Glottochronology is a method that was elaborated in the 1950s to evaluate language genealogies. This method estimated that every fourteen non-cognate pairs out of hundred-word lists corresponded to a thousand-year separation between languages (Swadesh, 1955). The word lists elaborated by Swadesh for this purpose were very influential in the discussion of borrowings but were also heavily criticized. Swadesh's word lists rely on the premises of the 'gap' hypothesis and on the assumption that some concepts are more likely to remain unaffected by language contact (e.g., body-parts, close kin, body-related activities, pronouns, interrogatives, and basic concepts for nature and geography). The limits of this method to evaluate language genealogical groupings and separations have long been identified and alternative methods are currently in use such as Bayesian linguistic phylogenetic dating and ASJP chronology (see Holman et al., 2011). In addition, researchers working in current

approaches no longer wish to exclude borrowings, but instead embrace the idea that language contact is ubiquitous and that it needs to be quantified and modelled (Holman et al., 2007). Following advances in contact linguistics, it has now become apparent that diverse contact settings will impact differently on language genealogies. For example, settings with patrilocal exogamy (where women move into their husband's group) can be expected to introduce structural and phonological/phonetic features as women are predominantly second language speakers of the husband's language. By contrast, trade is more likely to be associated to the introduction of lexical borrowings associated to artefacts.

4. CURRENT CONTRIBUTIONS AND RESEARCH

The Markedness Hypothesis

Elšík and Matras (2006) assessed the role of the Markedness Hypothesis in borrowing. According to the Markedness Hypothesis categories that show greater complexity and lower differentiation are considered as 'marked,' and those that show lower complexity but greater differentiation are considered 'unmarked.' Elšík and Matras (2006) analysed data from the Romani Morphosyntax Database and showed that markedness and borrowing do not correlate in a uniform way. In borrowing, replication of routine forms, which are more frequent in usage, are associated with 'unmarked' categories, accounting for the more frequent borrowing of categories such as lower numerals or nominative case, whereas reduction of processing load is associated with 'marked' categories, explaining the high borrowability of categories such as peripheral local relations, or markers of contrast, restriction or discontinuity.

The Specificity Hypothesis

Based on the analysis of a Dutch-Turkish corpus, Backus (2001) formulated the Specificity Hypothesis according to which words with specific meaning are more susceptible to borrowing than words with general meaning. Serigos (2017) tested the Specificity Hypothesis by examining English borrowings in a 24-million-word newspaper corpus of Argentine Spanish. She assessed specificity by examining the target word's environment and by considering that more specific words would have less variation in their surrounding context. The analysis confirmed the Specificity Hypothesis as it revealed that English borrowings were found in contexts that were similar and showed less variation in comparison to their native counterparts. Serigos further interpreted this finding as indirect evidence for the motivation of speakers to add nuances to the original concept.

Verschik and Backus (2012) also discuss specificity with respect to borrowing of bound morphemes. They argue that structural boundedness itself plays just a minor role in explaining constraints on borrowing, and suggest instead that bound morphemes are less likely to be borrowed because their meaning is non-specific. For example, in dialectal Russian spoken in Estonia, the bound Estonian intensifier marker -*gi* 'indeed' is borrowable on account to its function.

The Utilitarian motivation

Based on an analysis of a cross-linguistic sample, Matras (2007) proposed the Utilitarian motivation for borrowing. According to the Utilitarian motivation, more borrowable elements are associated more strongly with interaction settings in the donor language. This is also reflected in Matras's (2007) proposal that 'unique referents' such as terms for institutions, procedures, and concepts related to activities carried out in particular settings are more closely

associated with specific languages and therefore more susceptible to borrowing, as opposed to general/core vocabulary. Similarly, numerals in formal contexts, higher cardinal numerals, and days of week are found to be more susceptible to borrowing than numerals in informal contexts, lower cardinal numerals, and times of day. The domain-related hierarchy of numeral borrowing, for example, is confirmed in a study of the Tetun Dili language of East Timor in contact with Indonesian and Portuguese. Tetun speakers use three parallel numeral systems that reflect the interaction domains with which the respective languages are associated: alongside the Tetun numeral set, which is preferred when counting family members, there is wholesale borrowing of the sets of numerals from Indonesian, in connection with education, prices, and technical features, and from Portuguese, for dates and maths (Williams-Van and Hajek, 2018).

Support for a domain-related hierarchy also comes from the results of the typological project Loanwords in the World's Languages (Haspelmath and Tadmor, 2009). This database offers comparative data from 41 languages collected with a questionnaire that was filled by language specialists. The analysis of the sample shows that semantic fields related to social organisation such as the modern world, possession or commerce, religion, law, and household, are more susceptible to borrowing than semantic fields related to natural or physical surroundings such as physical world, body, emotions, perception, space, and kinship. Matras (2020) interprets these preferences as evidence for the tendency of bilinguals to generalize the use of some items in all contexts of use and interlocutors. Conversely, it is less likely for speakers to borrow items that are preferred in private settings where it is easier to maintain the demarcation between sub-components of the repertoire. Support for this view comes from the corpus study conducted by Chesley and Baayen (2010). The authors investigated borrowings in French in two corpora within a ten-year time difference and found

that borrowings were best preserved when they were shorter and more polysemous, and when they were used in a wide variety of contexts, including contexts that are not culturally specific.

Usage frequency

The usage frequency criterion has been widely discussed both in regard to the recipient language and the donor language. According to the usage-based account, the more frequent an item is in a given language, the more entrenched it becomes and the easier its activation is for a bilingual speaker.

According to Matras (2007), higher frequency, routine, and casualness of usage of words from one language discourages borrowing. This is the case of close kin and core local relations. At the same time, some studies provide compelling evidence for the role of frequency of the borrowed term in the donor language. Interestingly, frequency of an item in the donor language seems to affect the item's structural integration. For example, Backus (2003: 104) analysed a bilingual Turkish-Dutch spoken corpus from the Netherlands and noted that at least half of the Dutch compound nouns used in Turkish sentences were high-frequency Dutch items. Backus postulates the Unit Hypothesis to account for the fact that Dutch compound nouns are integrated as such in Turkish. Similarly, based on the analysis of a bilingual speech corpus and statistical frequencies in monolingual speech, Hakimov (2016) demonstrates that usage frequency in the donor language reliably predicts the insertion of inflected words in bilingual sentences. In particular, Hakimov found that nouns are more likely to retain plural inflection when they are more frequently used in plural (as opposed to singular form) in monolingual speech.

The impact of frequency of use on bilingual processing has been discussed in the study of Turkish verbs in a Romani variety spoken in Greece. First, in a corpus study of Turkish-Romani bilingual speech it was shown that verbs from Turkish are not integrated into Romani morphology but instead keep morphological marking intact (Adamou and Granqvist, 2015). The question in this case was whether these verbs were borrowings or codeswitching insertions given that speakers were bilinguals. To address this question, Adamou and Shen (2019) relied on additional behavioural data. Analysis of reaction times in sentence processing showed that Turkish-Romani bilingual speakers living in Greek Thrace process Turkish verbs differently depending on their likelihood to be used in Turkish or in Romani within their speech community. More specifically, well-established Turkish verbs were processed without cognitive cost, as in unilingual speech processing (speech in a single language), while Turkish verbs that retain a Romani counter-part were processed with cognitive costs as words that are typically associated to alternation between two languages. Adamou and Shen concluded that the data support usage-based approaches where frequency and interactional habits play a decisive role in processing.

Borrowability of morphology

Since Moravcsik (1978), it has been widely accepted that inflectional morphology is less borrowable than derivational morphology, and that within inflectional morphology, plural markers are borrowed more frequently than case markers or person agreement markers. To account for these tendencies, it has been noted that inflectional morphemes are more abstract, syntactically dependent on other elements (e.g., in agreement), and have a deictic or anaphoric function (e.g., in agreement) (see Matras, 2007; Gardani, 2008). Matras (2009) proposes the following hierarchy based on semantic transparency (also see Field, 2002 for the notion of transparency): derivation marker > classifier > plural marker > definiteness marker

> case marker. He argues that derivational procedures that mark agentivity, diminutives, and noun classifiers are highly transparent and are therefore susceptible to borrowing. Plurality is relatively transparent as it serves to categorise a noun in terms of quantity. Less transparent are the definite and indefinite articles that indicate a noun's pragmatic status in discourse. Finally, case markers are the least transparent as they rely on the semantic-syntactic role of nouns in a phrase.

The above-mentioned tendencies were largely confirmed in a recent cross-linguistic study dedicated to bound morphological affixes that draws on data from roughly 100 language pairs (Seifart, 2017). For instance, this overview confirms the previously noted prominence of nominal plurality markers among borrowed affixes. The special treatment of plurals as part of the lexical stem rather than as an inflectional ending in borrowed words is also apparent in Jerusalem Domari where speakers add the Domari inherited plural formation to Arabic-derived plural nouns: singular *zálame* 'man' (Arabic singular *zálame*), plural *zlām-é* (Arabic plural *zlām*) (Matras, 2009). Seifart (2017) also notes that verbal inflection is less borrowable than nominal inflection. A rare exception to this tendency comes from speakers of Éven (Tungusic) who borrowed subject agreement markers together with mood suffixes from Sakha (Yakut) (Pakendorf, 2019). Pakendorf considers that this rare outcome results from direct borrowing of the inflectional paradigms and notes three facilitating factors: the fact that both languages are agglutinative, that Sakha verb stems are easily segmented, and that the tense-aspect-mood markers need to combine with subject agreement suffixes. Indeed, the third factor might be seen as motivated by the drive toward what Matras (2009) calls 'fusion' of, in this case, operational markers of modality.

Social motivations for borrowing

Two recent studies based on the analysis of large spoken corpora investigated the role of social motivations for borrowing. In the first study, Zenner, Speelman, and Geeraerts (2015) analysed a 35-hour spoken corpus from a reality TV show from 46 speakers of Dutch to identify the use of English insertions. They found that English insertions into Dutch are preferred by younger, more educated males allowing them to highlight aspects of their identities. They also showed that, beyond the classic sociolinguistic variables such as age and gender, situational factors were crucial, as for example English swearwords were particularly frequent in contexts charged with negative emotions.

The second study demonstrates that borrowing does not always come from a socially dominant language. This is illustrated in Calude, Miller, and Pagel (2017), a study based on the analysis of a million-word spoken corpus of New Zealand English. The study confirms the high frequency of Māori borrowings despite the fact that Māori is a minority language. Rather, Māori has been going through a significant revitalization process since the 1970s that seems to be affecting its status in positive terms. More specifically, the authors found that social factors such as the ethnic identity of the speakers and of the audience accounted for the extent of Māori borrowing, where higher use of Māori borrowings signalled some sort of support for the Māori identity. Linguistic factors were further identified in the uses of both Māori and White/European New Zealand speakers of English: Māori borrowings were more likely to be used when they had a cultural rather than a core meaning, when they did not have a clear English equivalent, and when they were content words rather than function words.

Finally, Adamou et al. (2016) compared spoken corpora from bilingual speakers of four Slavic minority languages from settings of century-long contact with the majority languages

(specifically Greek, Italian, and German). Statistical analyses revealed patterns of borrowing within each community. An additional analysis including extralinguistic factors showed that the degree of lexical borrowing did not depend on the degree of bilingualism in the community or the duration of contact, but was rather conditioned by prescriptive attitudes supported by institutions.

5. MAIN RESEARCH METHODS

Early theoretical generalizations about borrowing relied on the study of language contact phenomena in individual language pairs. Dawkins (1916) for example is an important case study that drew attention to the extensive influence of Turkish on Anatolian Greek, including grammatical inflections, word order, and relative clause constructions. In addition to such qualitative studies, quantitative empirical evidence has become decisive to the development of theoretical claims in language contact. These come from structured cross-linguistic samples and corpus-based studies.

Since the 2000s, corpus-based approaches to language contact and the study of borrowing have become more central due to the growth of computerized corpora. Contact linguists work with written and spoken corpora: the former offer the possibility to explore large corpora, the latter offer direct access to the spontaneous productions of bilingual speakers and to the phonetic details that may be crucial in determining the status of an item as a borrowing. For well-described language pairs and written corpora, automatic language tag models are available using n-gram methods (calculating for example probabilities of character sequences to belong to a given language) and look-up methods (checking the languages' dictionaries) (see e.g., Guzman et al., 2016). In contrast, in the case of lesser-studied languages that lack

tools for automatic annotation, manual coding is required. Along with phonetic and morpheme annotation, researchers can tag the words depending on the languages involved: for example, L1, L2, multiple or ambiguous, mixed, other language, or unknown (Adamou, 2016). In the analysis of borrowing, in particular, some corpus linguists adopt a frequency-based approach and others a concept-based approach. In the former, statistical models are used with raw frequencies of borrowed items, whereas in the latter, borrowing use is relative to the use of a specific concept. Though the concept-based approach captures an important aspect of lexical borrowing, it restricts the study to a small set of words that have a more or less clear-cut semantic equivalent.

As discussed in this chapter, cross-linguistic samples have also been created for the study of borrowing. However, the study of borrowing in this perspective presents special challenges (see Matras and Sakel, 2007a). In addition to the need for the sample to reflect the diversity of languages and language families or areas, a comparative study of borrowing must take into account all the categories that are potentially affected by contact. Since the study of borrowing implies a diachronic perspective, information is required about the history of contact of the languages under study. However, reliable diachronic information is lacking with respect to the languages of entire regions. Moreover, as borrowing is sensitive to extralinguistic factors, these too need to be taken into account (e.g., duration and intensity of cultural contact, the roles and status of the participating languages, language attitudes, and the degree of institutional support enjoyed by the languages). Finally, information on contact phenomena is not always available in descriptive grammars and discussions of specific contact phenomena do not cover the entire set of contact phenomena present in a given language. Despite such difficulties, extant cross-linguistic samples of borrowing demonstrate

that it is possible to bring together language specialists and work with structured questionnaires that make systematic data comparison possible.

6. FUTURE DIRECTIONS

Some researchers consider translanguaging to be a useful lens through which to enhance our theoretical understanding of borrowing. However, as MacSwan (2017) notes, proponents of the translanguaging analyses need to offer a coherent account for the abundant empirical evidence in the literature and propose a solid frame that can explain how mental lexicons and grammars are organised in a bilingual mind. Usage-based approaches seem to offer such a coherent framework as they combine sociolinguistic, psycholinguistic, and linguistic aspects to account for both diachronic and synchronic processes in language contact (see Chapter 6).

The significant progress of electrophysiological and neuroimaging methods in the past decades can also be expected to impact on the study of borrowing as it has done in bilingual language processing. Behavioural methods have already been offering novel insights into the understanding of bilingualism and language contact phenomena (see Chapter 2). In connection with the codeswitching/borrowing discussion, for example, the behavioural data in Adamou and Shen (2019) open the possibility to consider borrowings as those items that would entail no processing costs as they are expected to be used in a specific language, whereas code-switching insertions would entail processing costs due to a higher surprise effect stemming from the choice of the language.

Lastly, Natural Language Processing methods are currently being applied to the development of new automated techniques, adapted to the specificities of bi/multilingual corpora (see Guzman et al., 2016 and Chapter 1). The analysis of more sizeable bi/multilingual corpora, representing a wider variety of contact settings and language pairs could offer a much-needed empirical basis on which to test extant theoretical models.

7. FURTHER READING

Matras, Y. (2020). *Language Contact*. 2nd ed. Cambridge: Cambridge University Press.

    This book offers an up-to-date introduction to the highly dynamic field of contact linguistics. It covers a wide-range of language contact aspects, ranging from bilingual child acquisition to typology.

Poplack, S. (2018) *Borrowing: Loanwords in the Speech Community and in the Grammar*. Oxford: Oxford University Press.

    This book is dedicated to the study of borrowing. Using the variationist approach, it relies on empirical data from a variety of language pairs. The book stresses in particular the key role of integration in the process of borrowing.

8. RELATED TOPICS

Codeswitching, mixed languages, the 4-M model, usage-based approaches, variationist methods, processing multilingual corpora

9. REFERENCES

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

Adamou, E., and Arvaniti, A. (2014). Greek Thrace Xoraxane Romane. *Journal of the International Phonetic Association*, 44, pp. 223–231.

Adamou, E., and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19, pp. 525–547.

Adamou, E., Breu, W., Scholze, L., and Shen, X. R. (2016). Borrowing and contact intensity: A corpus-driven approach from four Slavic minority languages. *Journal of Language Contact*, 9(3), pp. 515–544.

Adamou, E., and Shen, X. R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23, pp. 53–70.

Al-Hashmi, S. (2016). *The Phonetics and Phonology of Arabic Loanwords in Turkish: Residual Effects of Gutturals*. PhD. University of York.

Backus, A. (2001). The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In: R. Jacobson, ed., *Codeswitching Worldwide II*. Berlin and New York: Mouton de Gruyter, pp. 125–154.

Backus, A. (2003). Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41, pp. 83–132.

Blommaert, J., and Backus, A. (2013). Superdiverse repertoires and the individual. *Multilingualism and Multimodality*, pp. 11–32.

Calude, A. S., Miller, S., and Pagel, M. (2017). Modelling loanword success – a sociolinguistic quantitative study of Māori loanwords in New Zealand English. *Corpus*

Linguistics and Linguistic Theory, Online access only: https://doi.org/10.1515/cllt-2017-0010

Campbell, L. (1993). On proposed universals of grammatical borrowing. In: H. Aertsen and R. Jeffers, eds., *Historical linguistics 1989: Papers from the 9th International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 91–109.

Chesley, P., and Baayen, H. R. (2010). Predicting new words from newer words: Lexical borrowings in French. *Linguistics*, 48, pp. 1343–1374.

Dawkins, R. M. (1916). *Modern Greek in Asia Minor. A Study of Dialect of Silly, Cappadocia and Pharasa*. Cambridge: Cambridge University Press.

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

Field, F. W. (2002). *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins.

Gardani, F. (2008). *Borrowing of Inflectional Morphemes*. Frankfurt: Peter Lang.

Gómez-Rendón, J. (2007). Grammatical borrowing in Imbabura Quechua (Ecuador). In: Y. Matras and J. Sakel, eds., *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Walter de Gruyter, pp. 523–550.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, pp. 3–15.

Guzman, G. A., Serigos, J., Bullock, B. E., and Toribio, A. J. (2016). Simple tools for exploring variation in code-switching for linguists.
In: M. Diab, P. Fung, M. Ghoneim, J. Hirschberg, and T. Solorio, eds., *Proceedings of the Second Workshop on Computational Approaches to Code Switching*. Austin, TX: ACL, pp. 12–20.

Hakimov, N. (2016*).* Explaining variation in plural marking of single German noun insertions in Russian sentences. In: H. Beherens and S. Pfänder, eds., *Experience Counts: Frequency Effects in Language*. Berlin: de Gruyter, pp. 21–260.

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

Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26, pp. 210–231.

Heath, J. (1984). Language contact and language change. *Annual Review of Anthropology*, 13, pp. 367–384.

Heine, B., and Kuteva, T. (2005). *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.

Holman, E. W., Schulze, C., Stauffer, D., and Wichmann, S. (2007). On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguistic Typology*, 11, pp. 393–421.

Holman, E. W., Brown, C. H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., Sauppe, S., Jung, H., Bakker, D., Brown, P., Belyaev, O., Urban, M., Mailhammer, R., List, J.-M., and Egorov, D. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52, pp. 841–875.

Hyman, L. M. (1970). The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics*, 1, pp. 1–48.

Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychologica*, 128, pp. 416–430.

Li Wei. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39, pp. 261–261.

Loebell, H., and Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, pp. 791–824.

MacSwan, J. (2016). Code-switching in adulthood. In: E. Nicoladis and S. Montanari, eds., *Bilingualism across the Lifespan: Factors Moderating Language Proficiency*, 1st ed. Washington, DC: American Psychological Association, pp. 183–200.

MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, 54, pp. 167–201.

Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36, pp. 281–331.

Matras, Y. (2007). The borrowability of grammatical categories. In: Y. Matras and J. Sakel, eds., *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Walter de Gruyter, pp. 31–74.

Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.

Matras, Y. (2020). *Language Contact*. 2nd ed. Cambridge: Cambridge University Press.

Matras, Y., and Sakel, J. eds. (2007a). *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Walter de Gruyter.

Matras, Y., and Sakel, J. (2007b). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31, pp. 829–865.

Moravcsik, E. (1978). Universals of language contact. In: J. H. Greenberg, ed., *Universals of Human Language*. Stanford: Stanford University Press, pp. 94–122.

Myers-Scotton, C. (1993). *Social Motivations for Code-switching. Evidence from Africa*. Oxford: Oxford University Press.

Myers-Scotton, C., and Jake, J. L. (2014). Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition*, 17, pp. 511–525.

Myers-Scotton, C. M., and Jake, J. L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21, pp. 340–366.

Pakendorf, B. (2019). Direct copying of inflectional paradigms: Evidence from Lamunkhin Even. *Language*, 95, pp. e364–e380. https://doi.org/10.1353/lan.2019.0063

Peperkamp, S., and Dupoux, E. (2003). Reinterpreting loanword adaptations: the role of perception. In: M. J. Solé, D. Recasens, and J. Romero, eds., *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, pp. 367–370.

Poplack, S., and Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24, pp. 279–315.

Poplack, S., Sankoff, D., and Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, pp. 47–104.

Salmons, J. (1990). Bilingual discourse marking: code switching, borrowing, and convergence in some German-American dialects. *Linguistics*, 28, pp. 453–480.

Seifart, F. (2017). Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics*, 7, pp. 389–431.

Serigos, J. (2017). Using distributional semantics in loanword research: A concept-based approach to quantifying semantic specificity of Anglicisms in Spanish. *International Journal of Bilingualism*, 21, pp. 521–540.

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

Swadesh, M. (1955). Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics*, 21, pp. 121–137.

Thierry, G., and Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104, pp. 12530–12535.

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

van Hout, R., and Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, 6, pp. 39–62.

Verschik, A., and Backus, A. (2012). Copiability of (bound) morphology. In: L. Johanson and M. Robbeets, eds., *Copies versus Cognates in Bound Morphology*. Leiden: Brill, pp. 123–149.

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

Whitney, W. D. (1881). On mixture in language. *Transactions of the American Philological Association (1869-1896)*, 12, pp. 1-26.

Williams-Van Klinken, C., and Hajek, J. (2018). Mixing numeral systems in Timor-Leste. *NUSA: Linguistic Studies of Languages in and around Indonesia*, 64, pp. 65–94.

Wohlgemuth, J. (2009). *A typology of verbal borrowings*. Berlin: Mouton de Gruyter.

Yip, M. (2006). The symbiosis between perception and grammar in loanword phonology. *Lingua*, 116, 950–975.

Zenner, E., Speelman, D., and Geeraerts, D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism*, 19, pp. 333–346.