The Multilingual Lexicon: Modelling Selection and Control

ARTICLE *in* INTERNATIONAL JOURNAL OF MULTILINGUALISM · JANUARY 2004

DOI: 10.1080/14790710408668176

Kees De Bot
University of Groningen

The Multilingual Lexicon: Modelling Selection and Control

Kees de Bot
Department of Language and Communication, University of Groningen, The Netherlands

In this paper an overview of research on the multilingual lexicon is presented as the basis for a model for processing multiple languages. With respect to specific issues relating to the processing of more than two languages, it is suggested that there is no need to develop a specific model for such multilingual processing, but at the same time we are still very far from really understanding how languages interact within and between components of the model.

Keywords: psycholinguistic modelling, multilingualism, lexicon

Introduction

While in current linguistic theories and (some) psycholinguistic models the lexicon has lost its isolated position as a more or less independent operation component, in work on bilingualism and multilingualism the structure and workings of the lexicon are at the heart of current research. As pointed out in many of the contributions in Cenoz *et al*. (2001), few researchers in psycholinguistics have made the step from bilingualism to multilingualism and current models have limited themselves to the processing of two languages, despite the inevitable but rather vacuous disclaimer that 'the same holds for other languages'. To what extent current models have to be adapted is as yet unclear.

In our thinking on tri- and multilingualism, we should prevent ourselves from falling in the pitfalls of paradigm contests. There may be grounds to define trilingualism as a separate field within Applied Linguistics, and the recent flurry of activity including the present journal speaks to that. Whether this will be the direction for the future remains to be seen. Setting this research too far apart from what has been done in SLA may not be wise, even though it can be argued that quite some of the outcomes of that research might have been different had issues of the impact of more than two languages been taken into account.

Issues in the Bilingual Lexicon

Selective vs non-selective access

The first issue is how we access words from different languages in our lexicon. This question has replaced the more traditional one of whether words from different languages are stored in one or more containers. Such a localist perspective is not in line with the more functional approach in psycholinguistics, though it is coming back in more recent neuroimaging work on multilingualism.

1479-0718/04/01 015-16 $20.00/0

A distinction needs to be made between perception and production here. There is a huge amount of experimental research on perception, or to be more precise, on bilingual word recognition, and very little on production. There is an interesting paradox here: while we cannot control language production, we have to rely mainly on spontaneous speech data. In particular, codeswitching data are an enormously rich source of information that we have only begun to explore. At the same time we cannot manipulate real spontaneous speech which makes testing of hypotheses in experiments difficult, if not impossible. For language perception the opposite is true: we can very carefully manipulate stimuli to test all sorts of hypotheses, but we have no data on spontaneous listening. Interestingly, the same paradox re-emerges in research on neuroimaging: because of the neural noise generated by the muscles in our vocal tract, it is impossible to get reliable neuroimaging data on spoken language. Again manipulation is easy in perception, and we might hope that those techniques tell us at least something about normal listening: we can see things happening in the brain, while in experiments we only have external reactions like latencies to go by.

The central question most of the research is concerned with is language-specific versus non-specific access. In other words, when we are confronted with a word, e.g. in a lexical decision task, do we first access the lexicon from one language and then the next, or is there a parallel search through all languages, words not being organised primarily through language, but e.g. through frequency. In the past the language selective position was favoured (Gerard & Scarborough, 1989; MacNamara & Kushnir, 1971), but in the last two decades research has accumulated to support the non-selective view. A number of experimental paradigms have been used to study this question.

Neighbourhood effects in word recognition

Neighbourhood effects provide one set of evidence against selective access (see also Dijkstra & van Heuven, 2002). Here the argument goes as follows: in monolingual word recognition, reaction times for words that have many so-called neighbours, i.e. words that differ in one letter or sound from the target word, are longer than for words that have few neighbours. Examples for English are 'Band' (many neighbours) and 'Ilk' (few neighbours). Many neighbours means that there are many competing candidates for the target word and that makes it difficult to select the right word, leading to a slowing down of the process. In bilingual studies, this effect can be used to test the selective/non-selective issue: If access is selective, that is, specific for one language, neighbours from another language should have no impact on latencies. If however, access is non-selective, neighbours from both languages compete. By manipulating the number of neighbours in L1 (Dutch) and L2 (English), van Heuven *et al*. (1998) found that an increase in the number of Dutch neighbours leads to slower reaction times in an English lexical decision task, and this is supportive for non-selective access. Similar evidence is presented by Jared and Kroll (2001). Note that in few, if any studies, neighbours from other languages than the two focused on have been taken into account.

The processing of cognates and interlingual homophones/homographs

A whole branch of research has used two types of stimuli, cognates, words with similar form and meaning in two languages and words that are known as interlingual homophones to test the selective/non-selective hypotheses (Dijkstra *et al.*, 2000). Interlanguage homophones are words that sound similar in two languages but have a different meaning. Examples are 'Coin' in French and English or 'List' in English and Dutch. Similarity between the words of the two languages may vary somewhat and there is no clear definition of what 'similar' is, both in orthography and in phonology. Most researchers try to control the variability in this respect as much as possible.

Evidence from studies using cognates and interlingual homographs and homophones follows a similar logic to the neighbourhood studies. If access is language-selective, the fact that words are cognates of homographs should have no effect on reaction times. If access is non-selective, candidates from both languages will present themselves and this competition will again lead to longer reaction times. A large number of studies have been done on this, and there is overwhelming evidence to support the non-selective access hypothesis (see Dijkstra & van Heuven, 2002 for detailed overviews).

Cross-linguistic priming and repetition effects

A third set of studies used various forms of priming and repetition with lexical decision to test the selective/non-selective issue. In semantic priming experiments the target word is preceded by a word that is either semantically related or unrelated. That is, when the target word for lexical decision is 'Hospital', reaction times will be shorter when the target word is preceded by a related word like 'Nurse' than when it is preceded by an unrelated word like 'Bread'. In a repetition task the prime is the same as the target, and words that are repeated are typically recognised more quickly than words that are preceded by another, non-related prime. The methodology of this type of experimentation has become very sophisticated over the years with the use of a range of SOAs, i.e the time between prime and target, and of various forms of masking in order to manipulate the amount of conscious processing of the prime. These paradigms have also been used with bilinguals. Here, the selective access hypothesis would predict no effect of priming or repetition between languages (e.g. a comparison between the priming effect of 'Nurse' and its Dutch translation equivalent 'Verpleegster' versus a non-related prime on the English target 'Hospital').

Woutersen (1997) presents an overview of these studies and shows that for short SOAs (200–800 ms), interlingual priming does occur supporting the non-selective access hypothesis. For longer SOAs (2000+ ms), no interlingual effects were found, which may be the result of more or less conscious strategic processing.

Eye-tracking studies

Another source of information is data from eye-tracking studies. An example of such a study is Marian *et al.* (2002), who looked at English–Russian bilinguals. In their experiments they asked subjects to carry out tasks like 'Pick up the xxxx' and monitored their eye movements. In the experiment there were target objects, objects whose name in the other language showed overlap in form with the target name and control objects whose name had no overlap. In the experiments between-language and within-language competition was tested. In all conditions the between-language competition attracted more and longer gazes than the non-overlapping control objects. There was also a robust within-language competition effect. Effects were equally large for the two languages. The data are interpreted as evidence for simultaneous activation of both languages in the early phonetic stages of perception, even in a completely monolingual setting.

Brain-imaging studies

This is a very rapidly developing field, and bilingualism seems to be high on the agenda. It is beyond the scope of the present paper to even start to summarise the (often conflicting) data that have been produced over the last few years. This line of research is clearly still in a 'wild' phase in which dramatic but quite often non-replicable findings find their way into the research literature, and it will take some time before the lines are becoming clear (for an overview see Paradis, 2001).

An example worth mentioning is the study of Marian *et al.* (2002). In addition to the eye-tracking experiment reported above, they also carried out an fMRI study with six late Russian–English bilinguals who were presented with words and non-words in both their languages. The analyses showed that processing the second language involved more brain volume and that different centres of activation were involved in phonological processing for L1 and L2 in the Inferior Frontal Gyrus (Broca's area) but not in the Superior Temporal Gyrus (Wernicke's area). This study, based on a very small number of participants shows all the complexities of this type of research: variability within and between participants on both behavioural and neuro-anatomical levels, and a considerable gap between processing models and neuroimaging findings. That is, it is not at all clear how to interpret differences in brain volume involved or differences in centres of activation between languages.

Language production: Picture/word interference studies

With respect to language production, much less has been done. In a series of experiments, Hermans (Hermans *et al.*, 1998) used a complicated picture word interference task in which the task is to name a picture, while at different SOAs words with similar meaning or similar form characteristics are presented on the screen. He showed that for production also non-selective access is the better solution, although Costa and Caramazza (1999) present data which they interpret as support for a more selective access procedure.

Overall, the evidence in support of the non-selective access hypothesis is substantial, and much stronger than for the selective access hypothesis.

Speed of processing in L1, L2 and L3

From the research on bilingual processing it has become clear that in all sorts of tasks bilinguals are slower in processing their weaker language. Woutersen (1997) presents an overview of reaction times for L1 and L2 in

Table 1 Auditory lexical decision latencies (from Woutersen, 1997)

(1) Secondary school students	L1 Dutch	951
	L2 English	1060*
(2) University students of English	L1 Dutch	903
	L2 English	945 (*)
(3) Near-natives/members of staff	L1 Dutch L2 English	850 868 (ns)

bilinguals, and her own data on Dutch/English bilinguals also show this trend for the less proficient bilinguals in an auditory lexical decision task (Table 1).

It is not clear whether bilinguals are slower because the access of words simply takes more time or because the less proficient language calls for more control mechanisms which take their time.

In the 1970s Mägiste from Stockholm University carried out a number of experiments that included groups of bilinguals and trilinguals. The languages she looked at were German and Swedish in Sweden, and her trilinguals were typical migrants with a range of different L1s. Her aim was to assess to what extent bilinguals and monolinguals are slower in language processing, compared to a group of (more or less) matched monolinguals. She used a number of different tasks, partly decoding tasks (e.g. reading aloud printed words) and partly encoding tasks (picture naming and naming two digit numbers). Her data show that the trilingual group performed more slowly on most tasks as compared to the monolinguals and the bilinguals (apart from naming objects in Swedish). Though she does not give the exact figures, it can be inferred from the figures she presents that the trilingual group was some 200ms slower in the three remaining encoding tasks. In subsequent research, Mägiste (1986) showed that trilinguals performed worse on various parts of the Raven Matrices and were slower in bilingual Stroop tests, again in both languages German and Swedish. This research at least suggests that there may be a price to be paid for adding yet another language. Two hundred milliseconds may not sound very long, and indeed it doesn't take long for 200ms to pass, but 200ms is also the time we take on average to access words in our lexicon in language production. The timing of our language production and perception systems is extremely strict, and in that system 200 additional ms are significant and may indeed lead to processing problems. Whether they actually do, we don't know, because very little is known about the whole issue of timing constraints. Maybe our system is flexible enough to cope with such delays and we do compensate for this using other strategies in production.

Little is known about this. It may be that our processing mechanism has built up a processing rhythm based on the L1 and that therefore processing a language in which elements are accessed slower and therefore become available later is problematic because the subprocesses do not 'fit' any more. It may also be that there is a 'different drummer' in the system that adjusts the processing speed to the availability of elements. Given the accuracy with which the time domain can be measured in neuroimaging, it is likely to be one of the most interesting aspects to look at.

Can languages be 'switched off' in multilinguals?

The evidence in support of the non-selective access hypothesis seems to answer this question: if language access is non-selective, both languages must be active. Unfortunately things are more complicated than this. A crucial issue is to what extent informants were ever in what Grosjean (1999) has called the 'monolingual mode'. In his bilingual mode model, individuals who have a command of more than one language will be somewhere on a continuum from a completely monolingual mode to a completely bilingual or multilingual mode. In the experiments Dijkstra and his colleagues carried out, the informants appeared to be extremely sensitive to small variations in task demands and the composition of the word lists (see Dijstra & van Heuven, 2002). It could be argued that in the setting of most of the experiments in which psychology students served as subjects, the context was never completely monolingual, even though the stimuli used were only in one language. The simple fact that students tell their friends and co-students what they had to do in the experiments may have had an effect. Also the fact that most of the reading material during their study is in English and the fact that English has acquired a very prominent position in Dutch society in general may play a role here. Probably the Netherlands is not the right setting to test this hypothesis. Van Hell and Dijkstra (2001) present data from experiments with trilinguals (Dutch/English/French). Tasks used were timed word-associations and lexical decision in Dutch. There were four types of stimuli: non-words, Dutch/French cognates, Dutch/English cognates and non-cognates (i.e. not with Fr/Eng). According to the non-selective access hypothesis, even in processing the first and dominant language, there should be an effect of cognateness: they should be more easy to access. Psychology students (experiment 1) and students of French (experiment 3) were tested on their proficiency in all three languages. Both the association test and the lexical decision test showed a significant effect for English cognates over non-cognates, but not for French cognates in experiment 1, i.e. with students who had some French, but not a very high level of proficiency. In experiment 3, with highly proficient students of French, a cognate effect was found for both English and French as compared to non-cognates. This seems to suggest that words from another language need to have a level of activation that makes them accessible at a level that is not too different from words from the other languages. This is an important finding for our understanding of multiple language processing: in addition to being activated for a specific task, elements from a language need to have a default level of activation above a certain level. It is not the case that all languages, irrespective of level of proficiency, compete at the same time.

Inhibition of languages in switching

Recently a series of studies using different paradigms have looked at code-switching (CS) in experimental settings. A highly relevant publication for this discussion is Meuter and Allport's (M&A) article (1999) on language switching costs in experimental settings. Most research on CS makes use of spontaneous speech data, but M&A continued a much smaller line of experimental research which started with the work of Macnamara in the 1960s (Macnamara, 1967).

M&A's main point is that in CS processes are at work that are similar to control mechanisms in other, also non-language, task-switching activities. Their argument is that using the L2 requires active inhibition of the stronger L1. When switching from L2 to L1, this deactivation of L1 persists, making access of the L1-items more difficult than switching to L2, because L2 is much less deactivated, and activating it takes less effort. This is in a way counterintuitive, because switching to the stronger language would seem to be easier.

Their data also show that language subsets are activated/inhibited as complete sets, not on the basis of individual elements. This supports the idea of languages as sets that can be activated and inhibited as proposed in various models for bilingual processing (cf. Paradis, 1987, 2001). The switch effect was found only for the first switched item, by the next item latencies were the same as any following item from the same language. Interestingly the effect is not affected by training the weaker language (along with the stronger language, which of course should profit less from such training). Such training could lead to a smaller dissymmetry between the items from the two languages, but this was not borne out by the data. Apparently the activation/inhibition effects reflect a long history of use and practice of the language that cannot be easily overruled by short training sessions.

Following similar lines, Jackson *et al.* (2001) carried out forced switching experiments in which participants with different L2 learning backgrounds had to name numerals either in L1 or L2 in a mixed list in a predictable sequence of L1/L1/L2/L2/L1 etc. In order to prevent interference from muscular activity on neuroimaging data, two conditions were compared: fast (on presentation of numeral) or slow (after presentation), which yielded behavioural data (RTs) and neuroimaging data (ERPs). The experiment was similar to the one reported on by Meuter and Allport (1999). The RT data show that more inhibition is needed to switch to L1 than to L2 as compared to non-switch items. The same neural substrate seems to be involved in simple Go/no-Go tasks, which suggests that 'response suppression and response switching share a similar neural substrate' (1999: 173).

The Meuter and Allport and the Jackson *et al.* studies are interesting examples of a development we can expect more of in the near future: behavioural on-line measures in combination with neuroimaging data which will allow us to relate behaviour to activity in the brain both with respect to the timing and to the location of the activity.

But it also takes us back to the early days of bilingual research when Kolers (1963) suggested the existence of a language switch. Though appearing in a different guise, the mechanism and its neural substrate seem to be doing exactly what Kolers said about the language switch.

Modelling the Multilingual Lexicon

From the research presented, a number of lines can be deduced that should play a role in our thinking about the multilingual lexicon.

- Access to words in the lexicon is non-selective, i.e. words from more than one language compete for activation both in production and perception, but a – still to be defined – minimal level of proficiency/activation is needed to have words from a language play a role in the selection process, i.e. their default level of activation should be high enough to make them competitive.

- Non-selective access does not mean that words from whatever language have an equal chance of being selected, since languages as sets can be activated and inhibited. Language that are used often and have therefore a high default level of activation are difficult to suppress or inhibit, but once deactivated are also more difficult to activate.
- Shared forms at the phonological level tend to co-activate elements from different languages.
- Level of proficiency is reflected in latencies in various experimental tasks: lower levels come with longer latencies in on-line experimentation and higher error rates in judgement tasks.

These findings should be taken into account when developing a model of the multilingual language user. One crucial aspect has not been dealt with so far. Most bilinguals are actually multilinguals, though this is often not taken into account in the design of the studies. For instance, a lot of research has been done with Dutch–English bilinguals, but apart from Van Hell and Dijkstra (2001), the other languages Dutch university students typically have, such as French and German, are not taken into account. The model to be developed should be able to deal with more than two languages, and explain how the different languages interact in processing.

We will take Levelt's 'Speaking'-model in the 1993 version as a starting point. The main reasons are that the Levelt model in its various versions has a very strong empirical basis, and that in this particular version the perception side is also taken into account though to a lesser extent than production. Extensive discussions of the bilingual versions of the Levelt model can be found in De Bot (1992), Poulisse (1997), Pienemann (1998). Here only the main lines will be discussed. Figure 1 presents a highly simplified summary of the model.

At the highest level the conceptual information is represented. There is quite some discussion on how things are organised here. Concepts can be seen as bundles of semantic/conceptual features. Concepts that overlap in meaning share features, and accordingly the activation of a concept implies not only the activation of the features, but also of other concepts that are linked to those features. Semantic priming effects are based on this sharing of features. 'Language' may be one of the features involved. Lemmas are activated by matching the meaning components in the concepts and the lemmas. Words from different languages will share semantic features thus providing a bridge between languages through the conceptual system.

In the next step, lemmas will activate syntactic procedures. Now how these procedures are stored and activated is not at all clear. For our present purposes a major question is whether there is a set of, for example, English/German/Punjabi procedures in fixed sets that more or less define the grammar of a particular language. It seems more likely and more efficient to assume that we have a fairly large system containing all syntactic procedures that we have ever acquired which is governed by a universal processing

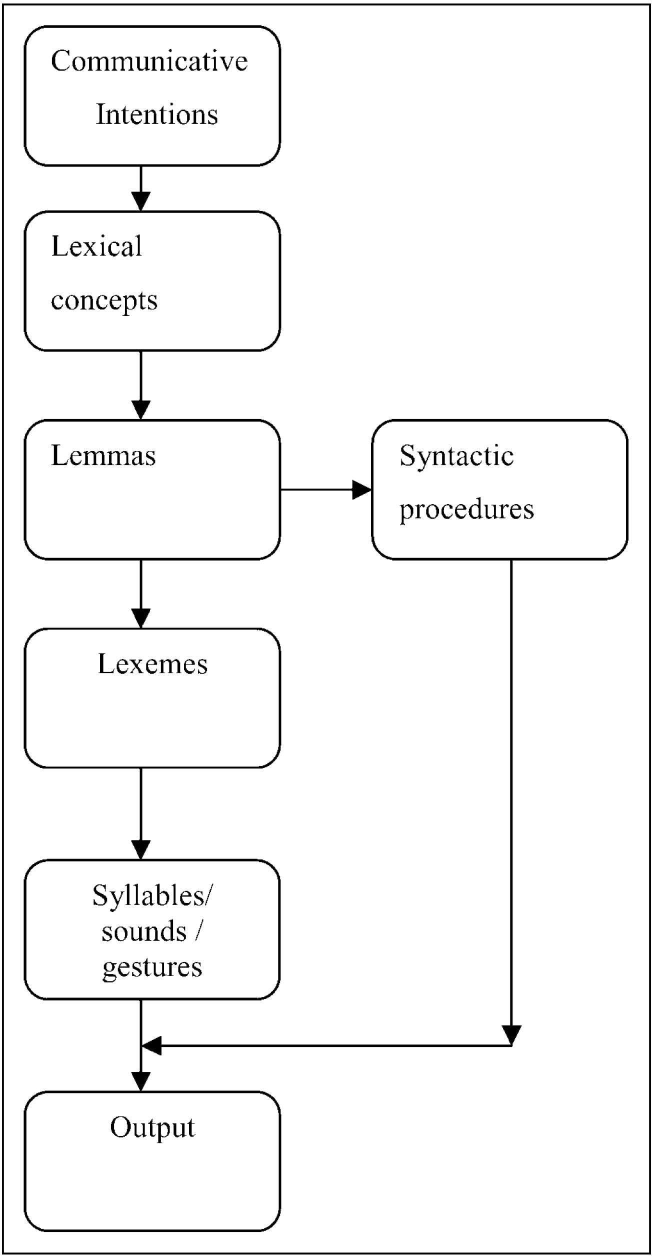

Figure 1 Schematic presentation of Levelt's Speaking Model

system that limits the combinations of procedures mainly for processing reasons. For a particular language a limited subset of procedures is used on a regular basis, and activating one of those procedures means that the other procedures in that subset will also increase in activation. Activating a particular lemma, and accordingly the syntactic procedures triggered by it, means activating a procedure that is shared by various languages. For example, inversion of subject and verb in subsidiary clauses is a procedure shared by languages such as German and Dutch, but not English. Analogous to what happens at the conceptual level, sharing syntactic procedures leads to activation of all the subsets that share that procedure.

It should be added that there seems to be no research on cross-linguistic syntactic priming, so this part is basically speculation with no data to support it.

Moving on to the next phase, the lemma will (parallel to the activation of the syntactic procedures) look for its lexeme. Activating the lexeme starts a complex process of word formation procedures.

Levelt (1993) has proposed that the syllabary is the system containing information on word forms, syllables being the most efficient way to build words. This means that there is a large set of syllables stored, and again a subset of that larger set will be active when we speak a specific language: the syllable inventory of that language. Again syllables (or sounds or gestures) are shared by subsets for different languages and activation of a syllable from a subset will activate the other syllables in that subset. Shared syllables are again a bridge between subsets and accordingly between languages. The well-attested between-language phonological priming effects (see Woutersen, 1997 for an overview) are based on this. It is quite likely that for language learners syllables have to be constructed anew every time, using sounds or combinations from specific languages, and only later the more efficient syllabary-based production emerges. The fact that in languages we are fluent in, individual sounds are encapsulated in syllables and therefore less easily adapted, may be one of the reasons why it is so difficult to reach a native-like pronunciation.

Levels of activation

The models we are using, be it Levelt's models and its multilingual variants or Dijksta and Van Heuven's BIA model, all make use of the concept of level of activation of languages as sets. Languages differ in level of activation, and this level of activation will depend on amount of contact and use, level of proficiency reached, maybe method of instruction, age of acquisition and many more variables. We assume that at a given moment in time, each language has its default level of activation. The much used L1 will have a high level of activation, while a language learned in a distant past for a few weeks will have a very low level. Activating a language means enhancing its level of activation and maybe – but here we hit one of the hotspots in current research on bilingual word recognition – lowering the level of activation of other languages. The fact that languages differ with respect to their default level of activation means that we need some system to make it possible to speak a language with a lower level of activation to start with. Without such a system, the stronger language always wins.

Activation, and in particular inhibition will never be a like an on/off switch. It is more like holding down ping-pong balls in a bucket filled with water: With your hands you can hold down most of the balls, but occasionally one or two will escape and jump to the surface. Likewise, complete suppression of a language, in particular one with a high level of activation may be impossible. Therefore inference of the stronger language into the weaker language is more likely than inference from the weaker into the stronger. There is ample evidence for this from research on, for example, Dutch migrants in the US and Australia, who when interviewed in Dutch had all sorts of problems suppressing their English, the language they usually spoke, but hardly any suppressing their Dutch when speaking English (Ammerlaan, 1996; De Bot & Clyne, 1994).

For trilingualism an intriguing problem arises here: the common observation of the impact of the second language on the third. In a simple model one would expect a much stronger effect of the L1 on the L3 than of the L2. Few controlled experiments have been done on this as far as I know, but evidence from Hammarberg (2001) and Dewaele (1998) suggests that the L2 has more of an impact than its relative level of activation would predict. As Clyne (1997) has observed, convergence between languages is likely to play a role here. No research with typologically sufficiently different languages in a controlled experiment is available to substantiate this claim, therefore further speculation on this is pointless. Another explanation could be that because the first language is used more, it forms a stronger network which accordingly can be deactivated as a whole more easily than the more loosely organised second and third languages.

The whole issue of how we have access to languages that differ in level of activation is linked to theoretical discussions on access of competing information more generally. What is the impact of our intention to speak the weaker language, does it lead to additional activation of that language up to the point where it beats the stronger language, or is the stronger language suppressed to a level below that of the weaker language, or both? Dijkstra and Van Heuven (1998) have recently developed a model for bilingual word recognition in which inhibition no longer plays a role, but whether that model is equally valid for everyday language production as for controlled experiments remains to be seen.

So far, the discussion has been restricted to a steady state model, i.e. a model of the language user at a given moment in time. What we need is a model that does that, but also can account for changes over time. Therefore it may be useful to link this to Herdina and Jessner's *Dynamic Model of Multilingualism* (2002) which focuses on 'the dynamics of the processes involved in progression and regression on an individual level and the complex interdependencies between the factors involved in the language acquisition process' (2002: 25). In this model the emphasis is on individual differences and variation, and on languages as psycholinguistic systems. From this perspective, languages are highly individual, idiosyncratic subsets of elements at different levels, with no real boundaries to keep them apart, and always in a state of flux.

The model can also accommodate the kind of variation in L3 speakers mentioned by Dentler (2000) in her study on Swedish/English/German trilinguals, who in a writing task appear to be mixing correct and incorrect forms. This reflects the varying levels of availability of elements from the various languages even within short ranges of time.

In the bilingual literature three models are dominant: the Bilingual Interactive Activation model of Dijkstra and his colleagues (Dijkstra & Van Heuven, 1998), Kroll's Revised Hierarchical model (Kroll, 1993; Kroll & Stewart, 1994) and Grosjean's Language Mode model (Grosjean, 1997). While the first two models focus on word recognition and processing at the conceptual/lexical level, the third model has a more general aim of explaining how levels of activation of languages are reflected in bilinguals' position on a monolingual–bilingual continuum. In the model presented here, elements from different models are included, but it has a different goal: to try to relate proficiency and levels of activation to ongoing language processing, in particular with respect to language production. Again, it is a very simplified model of an extremely complex process, and in contrast to models like the BIA, it does not have the specificity needed to run simulations, though it may be developed to that point.

The Multilingual Processing Model

A schematic overview of the model is given in Figure 2. The basic assumption behind the model is that there are three stores with information: conceptual features, syntactic procedures and form elements (sounds, syllables or gestures). Within each of these stores, there are language-specific subsets. These subsets show overlap reflecting the cognateness of the languages involved. The language node controls the various processing components with respect to the language to be used. The intention to use a specific language originates at the conceptual/communicative intention level and is relayed to both the system generating lexical concepts and the language node. For the subsequent components the information on the language to be used now comes from two sources: through the lexical concepts and directly from the language node. In earlier proposals it was suggested that all information on language choice needs to be included in the lexical concepts, but some aspects (such as deliberately speaking with a foreign accent) seem difficult to control in such a way. An external language node system can more locally control language choice. When a particular language is called for, the language node will inform all relevant components, that is those components in which syntactic or form information needs to be selected, about the subset to be activated. This higher activation will lead to the selection of elements from the right language. Crucially, information about the activation of subsets (either through the lexical concepts or through the language node) will be exchanged between subsets of the various languages. In other words when the subset of syntactic procedures of language A is activated, this information will be relayed to both the language node (which will then forward that information again to the rest of the system), and to the subset of form-related elements from language A. So activation of a part of a language will activate elements from that same language at other levels.

At the same time, because subsets overlap, elements that are shared by more than one language may also activate the other subset (or subsets) they are part of. The language node conveys information about language selection both from the conceptual level to lower level components and between components at these lower levels. It accumulates information of the state of activation of various languages and acts in that sense as a monitoring device which

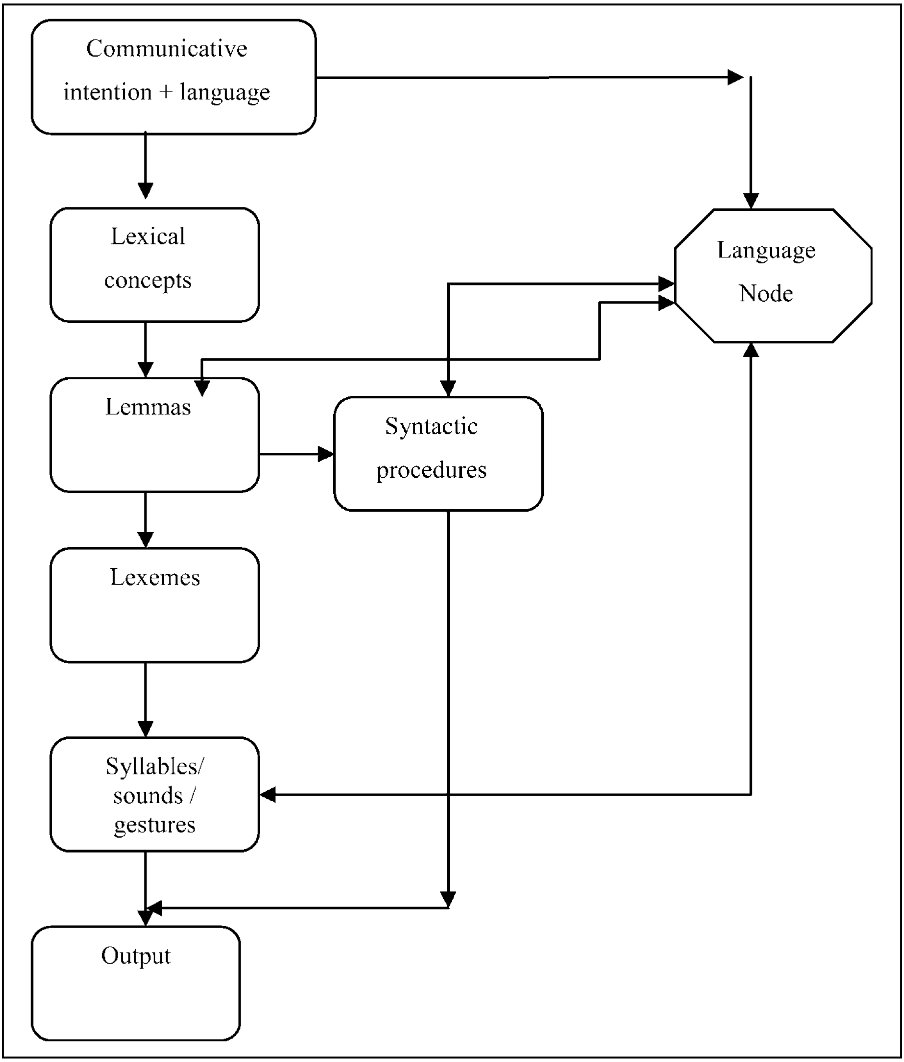

Figure 2 The multilingual processing model

compares the intended language with the language actually used. Evidence for this comes among other things from the between-language form priming found by Hermans *et al*. (1998). Other evidence comes from findings on triggering in CS as proposed by Clyne (1980) and tested by Broersma and De Bot (in preparation).

It should be clear that in this model 'a language' is a highly idiosyncratic constantly changing collection of elements. It has little to do with what a language is according to grammar books and dictionaries, and individuals may vary extensively with respect to their respective versions of the language. It may be that there are universals on both the syntactical and the form level which limit the possible combinations of elements in sets.

The current main issues in bilingual processing research, can be accommodated in the present model. Since the version of the model given is geared towards language production, some of the issues in perception, in particular the specifics of word recognition process are not represented in this model. Non-selective access can be accounted for by the idea of having subsets, which are more highly activated than the rest of the elements. In selection the elements from a given language/subset are preferred, but the elements from other languages are not completely blocked out.

The data showing that languages cannot really be switched off completely are in line with the idea of various levels of activation of languages, but again, as mentioned earlier, there are grounds for assuming a threshold for elements in competition.

How speed of processing can be implemented is not really clear. There is likely to be an interaction between level of activation and speed of processing in the sense that elements that become available earlier have a higher chance of being selected than elements that take more time, because the time constraints in language processing are tight.

Multiple languages pose no problem to the model. Note that the model presented does not make a real distinction between bilingualism, trilingualism and multilingualism, since the same principles apply for all situations. They also hold for different dialects, styles and registers within languages, which may form subsets in similar ways that languages do.

While various types of CS could already be accounted for in earlier bilingual versions of the Levelt model, the addition of the language node makes it easier to explain how switches at various 'low' levels in the system can be explained: the direct control by the language node puts less of a burden on the encapsulated information in the lexical concepts.

In this article a model for processing multiple languages is presented. With respect to specific issues relating to the processing of more than two languages, it is suggested that there is no need to develop a specific model for such multilingual processing, but at the same time we are still very far from really understanding how languages interact within and between components of the model. In the future the model has to be developed further and tested against data from different languages and from speakers of multiple languages and with different levels of proficiency. In an even more remote future, the model may be related to findings from neuroimaging research.

Acknowledgements

The author is indebted to Mirjam Broersma and the Editors for comments on earlier versions of this article.

Correspondence

Any correspondence should be directed to Kees de Bot, Department of Applied Linguistics, University of Groningen, Oude Kijk in 't Jatstraat 26, 9700AO Groningen, The Netherlands (c.l.j.de.bot@let.nig.nl).

References

Ammerlaan, T. (1996) *'You get a bit wobbly...' Exploring Bilingual Retrieval Processes in the Context of First Language Attrition*. University of Nijmegen.

Broersma, M. and de Bot, K. (in preparation) An alternative for the triggering hypothesis.

Cenoz, J., Hufeisen, B. and Jessner, U. (2001) *Cross-linguistic Influence in Third Language Acquisition*. Clevedon: Multilingual Matters.

Clyne, M. (1980) Triggering and language processing. *Canadian Journal of Psychology* 34 (4), 400–406.

Clyne, M. (1997) Some of the things trilinguals do. *International Journal of Bilingualism* 1 (2), 95–116.

Costa, A. and Caramazza, A. (1999) Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition* 2, 231–244.

de Bot, K. (1992) A bilingual production model: Levelt's Speaking model adapted. *Applied Linguistics* 13 (1), 1–24.

de Bot, K. and Clyne, M. (1994) A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development* 15 (1), 17–28.

Dentler, S. (2000) Deutsch und Englisch – das gibt immer Krieg. In S. Dentler, B. Hufeisen and B. Lindemann (eds) *Tertiär- und Drittsprachen. Projekte und empirische Untersuchungen* (pp. 77–97). Tübingen: Stauffenburg.

Dewaele, J-M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics* 19 (4), 471–490.

Dijkstra, T., Grainger, J. and van Heuven, W. (1999) Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41, 496–518.

Dijkstra, T., Timmermans, M. and Schriefers, H. (2000) Cross-language effects on bilingual homograph recognition. *Journal of Memory and Language* 42, 445–464.

Dijkstra, T. and van Heuven, V. (1998) The BIA-model and bilingual word recognition. In J. Grainger and A. Jacobs (eds) *Localist Connectionist Approaches to Human Cognition* (pp. 189–225) Hillsdale, NJ: Erlbaum.

Dijkstra, T. and van Heuven, W. (2002) The architecture of the bilingual word recognition system. *Bilingualism: Language and Cognition* 5, 175–197.

Gerard, L. and Scarborough, D. (1989) Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15 (2), 305–315.

Grosjean, F. (1997) Processing mixed language: Issues, findings and models. In A. de Groot and J. Kroll, (eds) *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 225–254). Hillsdale, NJ: Erlbaum.

Grosjean, F. (1999) The bilingual's language modes. In J. Nicol (ed.) *One Mind, Two Languages: Bilingual Language Processing* (pp. 1–25). Oxford: Blackwell.

Hammarberg, B. (2001) Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, and U. Jessner (eds) *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 21–41). Clevedon: Multilingual Matters.

Herdina, P. and Jessner, U. (2002) *A Dynamic Model of Multilingualism. Perspectives of Change in Psycholinguistics*. Clevedon: Multilingual Matters.

Hermans, D., Bongaerts, T., de Bot, K. and Schreuder, R. (1998) Producing words in a foreign language: Can speakers prevent interference from their first language. *Bilingualism: Language and Cognition* 1, 213–229.

Jackson, G., Swainson, R., Cunnington, R. and Jackson, S. (2001) ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition* 4, 169–178.

Jared, D. and Kroll, J. (2001) Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language* 44, 2–31.

Kolers, P. (1963) Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior* 2, 291–300.
Kroll, J. (1993) Accessing conceptual representations for words in a second language. In R. Schreuder and B. Weltens (eds) *The Bilingual Lexicon* (pp. 53–82). Amsterdam: Benjamins.
Kroll, J. and Stewart, E. (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33, 149–174.
Levelt, W. (1993) Language use in normal speakers and its disorders. In G. Blanken, E. Dittman, H. Grimm, J. Marshall, and C. Wallesch (eds) *Linguistic Disorders and Pathologies. An International Handbook* (pp. 1–15). Berlin: Walter de Gruyter.
Macnamara, J. (1967) The bilingual's linguistic performance: A psychological overview. *Journal of Social Issues* 23, 59–77.
Macnamara, J. and Kushnir, S. (1971) Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior* 10, 480–487.
Mägiste, E. (1986) Selected issues in second and third language learning. In J. Vaid (ed.) *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives* (pp. 97–121). Hillsdale, NJ: Erlbaum.
Marian, V., Spivey, M. and Hirsch, J. (2002) Shared and separate systems in bilingual language processing: Converging evidence from eye-tracking and brain imaging. *Brain and Language* 86 (1), 70–82.
Meuter, R. and Allport, A. (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40, 25–40.
Paradis, M. (1987) *The Assessment of Bilingual Aphasia*. Hillsdale, NJ: Erlbaum.
Paradis, M. (2001) Bilingual and polyglot aphasia. In R. Berndt (ed.) *Handbook of Neuropsychology* (2nd edn) (pp. 69–91). Amsterdam: Elsevier.
Pienemann, M. (1998) *Language Processing and Second Language Development. Processability Theory*. Amsterdam: Benjamins.
Poulisse, N. (1997) Language production in bilinguals. In A. de Groot, and J. Kroll (eds) *Tutorials in Bilingualism* (pp. 201–224). Mahwah, NJ: Lawrence Erlbaum.
van Hell, J. and Dijkstra, T. (2001) Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review* 9, 780–789.
van Heuven, W., Dijkstra, T. and Grainger, J. (1998) Orthographic neigborhood effects in bilingual word recognition. *Journal of Memory and Language* 39, 468–483.
Woutersen, M. (1997) *The Organisation of the Bilectal Lexicon*. Nijmegen: University of Nijmegen.