Making sense of prepositions:
The role of frequency and similarity in the acquisition of L2 prepositions

Wander Lowie and Marjolijn Verspoor
University of Groningen

Introduction

The most obvious difference between the acquisition of a first language and the acquisition of a second language is that in first language (L1) acquisition concepts and words are acquired simultaneously, whereas in second language (L2) acquisition the learner already has a fully developed set of conceptual representations. Also in learning the lexical items of a second language, the learner can be assumed to make use of the L1 knowledge already acquired. It is therefore not surprising that recent debates on the bilingual mental lexicon centre on the role of the first language and the development of this role over time. In this paper we will contribute to this discussion, focussing on the Dutch and English prepositions. Starting from an interactive activation model of the mental lexicon, we investigated the comparative effect of variables representing L1 influence and L2 influence, operationalised as prepositions categorized according to the dimensions of similarity and frequency.

Similarity and frequency in the bilingual mental lexicon

All recent (psycholinguistic) models of the mental lexicon tend to agree on the general contents of an item in the lexicon. Each item will refer to at least three separate units of information, semantics, syntax and phonology/orthography, which are divided between lemmas and lexemes. Following Levelt (1989), the lemmas are referred to as abstract units comprising the syntactic and semantic information, whereas the lexemes refer to the orthographic and phonological information associated with a lexical item, as illustrated in Figure 1.
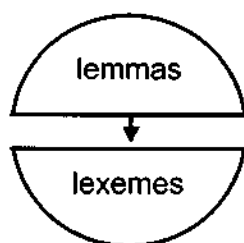

Figure 1 The basic components of a lexical item

In interactive activation models of the bilingual mental lexicon (cf Lowie, 1998, 2000; Schreuder&Baayen, 1995) the lexical item is represented accordingly. Here, each lexical representation comprises a lemma node, which is the central node linking the semantic-pragmatic information, the syntactic properties and the orthographic-phonological information (the lexeme). The model takes a compositional view on the relation between the semantic contents of the lexical item and the conceptual representations associated with it, in which the latter must be seen as the different aspects of semantic content of a word. Through a mechanism of activation and inhibition, the level of "resting" activation is primarily determined by the frequency of a lexical item. A frequently occurring word will have a higher level of activation than a word that scarcely used. Figure 2 schematically

represents a simplified representation of an item in the mental lexicon. Although different lemmas may share conceptual representations, no two lemmas in the mental lexicon can refer to a fully identical set of conceptual representations. In other words, this model allows partially overlapping word meanings, but it will not allow pure synonyms, as these would entail fully redundant items in the lexicon.
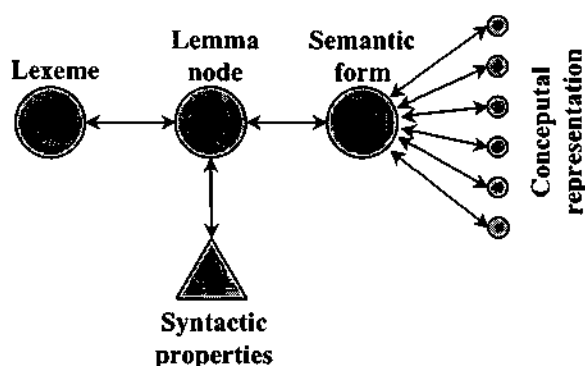

Figure 2. A simplified representation of a lexical item in the mental lexicon.

This image of lexical item in the mental lexicon can be adjusted to the *bilingual* lexicon by assuming an additional source of information linked to the lemma node referring to the language a lexical item is associated with. The items associated with a particular language can be regarded as a subset of the lexicon (cf. Lowie, 1998; Woutersen, 1997)

A question that is relevant to the bilingual mental lexicon is whether and to what extent L2 learners make use of the lexical knowledge from their first language in the acquisition and use of the second language. As the adult L2 learner possesses a fully developed lexicon, it makes sense to assume that an L2 learner will make use of the knowledge already required. A question that is central in current debates on the bilingual mental lexicon is whether L2 words have direct links with conceptual memory or are accessed through L1 lemmas present in the lexicon. Recently, Nan Jiang (2000) argued that the role of the first language differs in three stages of development. In the first stage, L2 forms are mapped onto existing (L1) meanings. At this stage, an 'empty' L2 lemma is created that is linked to a L1 lemma: the L2 lexical item only has the formal characteristics and full equivalence to an L1 lexical item is assumed. At the second stage, the information of an existing L1 lemma is copied onto the L2 lemma: this is the situation where the L1 lemma mediates L2 word processing. Only at the third stage will there be a direct link between the conceptual representation and the L2 lemma. This model sketches a picture similar to the one proposed by, for instance, Krol (1993), in which lexical items in L1 and L2 are connected:
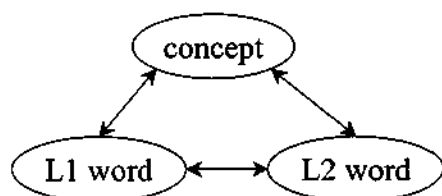
Although this model conveniently explains what Jiang calls "lexical fossilization", it cannot account for the fact that lexical items in L1 and L2 hardly ever fully overlap in meaning. In a model that takes compositional meanings as a starting point, this can be accounted

for much more easily. By referring to the activation metaphor, it is no longer necessary to distinguish between different ways of lexical organisation; activation models hypothesise that all individual lexical entries are stored identically, but that major differences between the entries can be expected based on their frequency, which is expressed by their relative level of activation. L1 entries are never directly linked to L2 entries, but information that is shared between the languages will result in activation feedback flowing to the lemma nodes concerned. In other words, L1 and L2 entries can never be lexically mediated, but are always conceptually mediated to a degree that is dependent on the relative activation of the conceptual representations, the lemma nodes and the lexemes. Similar to the way in which partially overlapping meanings in the monolingual mental lexicon can be accounted for in this way, this model can also account for overlapping meanings between L1 and L2. Figure 3 exemplifies the partial overlap between a Dutch and an English item in the bilingual mental lexicon.
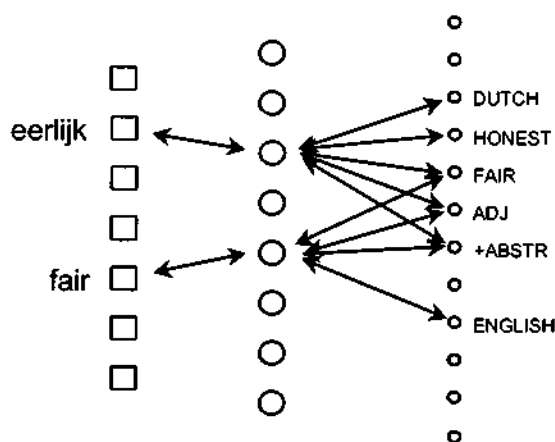

Figure 3 An example of partial overlap between lexical entries in L1 and L2 In this simplified representation, the different units of information associated with a lemma (semantic, syntactic, language) have been collapsed

The same framework can also be used to account for the *development* of the bilingual mental lexicon. At initial stages of L2 acquisition, a full overlap may be assumed between the conceptual representations of the L1 lemma and the L2 lemma. Gradually, the differences between the L1 and the L2 lemma will be acquired, which may eventually lead to a 'native-like' lexical representation. This process can entirely be based on positive evidence and is guided by the same principle that is at work in L1 acquisition: contrast (cf. Clark, 1993). When the learner encounters a new L2 word, this may lead to the partial restructuring of the semantic form of existing concepts by adding or deleting the match with some of the conceptual representations. This process is exemplified in Figure 4[1]. At some early stage of acquisition (t₁), the Dutch learner of English will assume full overlap between *between* and *among*, since in Dutch no conceptual distinction is made between these words[2]. Subsequently, the principle of contrast will ensure that the learner will not accept two lemmas to be fully identical, leading to the discovery of the semantic differences between *between* and *among*. This will subsequently lead to restructuring of the semantic form of *between* and the creation of a new lexical item *among*. The ultimate result of the acquisition process can be a "balanced" bilingual lexicon in which all semantic forms of all

---

[1] Analogous to an example worked out in Schreuder&Baayen (1995)

[2] The question is whether there is no conceptual distinction between these two words in Dutch or whether the same form happens to be used for the two different concepts (homonymy) This is an interesting side issue that goes beyond the scope of this paper.

lemmas have been fully specified. However, cases where this happens for all lexical entries in both languages will be highly exceptional, as most bilinguals will not be fully "balanced". The additional advantage of this approach is that it is no longer necessary to assume the same stage of development for entire language subsets. While some L2 lexical items may be fully developed, including all semantic and syntactic regularities and restrictions, other items may be found in different stages of acquisition.
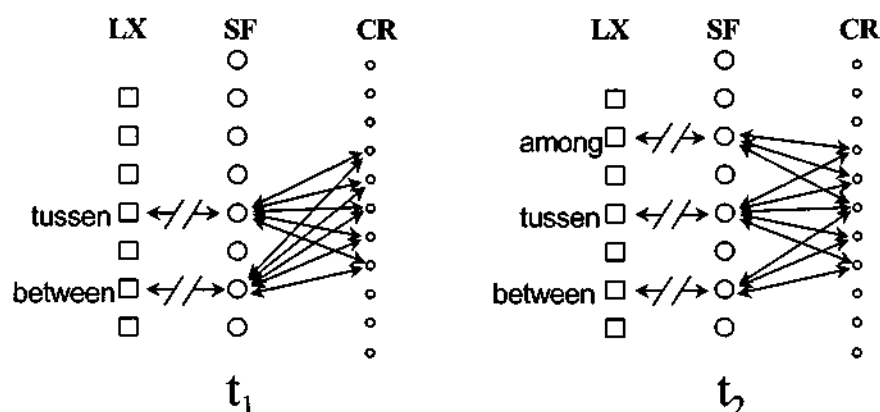

Figure 4 Simplified representation of two time slices in the process of acquiring the new L2 concept "among" In this figure, the lemma nodes have been left out

After this elaboration of the interactive activation model of the bilingual mental lexicon, let us now return to our main question: what is the role of the first language in the acquisition of lexical items in L2? In the interactive activation model, cross-linguistic similarity can be expected to affect the acquisition of L2 lexical items at two levels. First, it can be expected to occur on the left-hand side of the model, at the end of the phonological / orthographic representations. Orthographic and phonological similarity to L1 lexical items may affect the acquisition of L2 lexical items. This effect may be facilitating in the case of cognates, but can be confusing when formal similarity does not coincide with semantic similarity (Lowie, 1991). Second, cross-linguistic influence will be related to the degree of semantic overlap between lexical items in different languages. This effect, *Translation Equivalence* (Lowie, 1998) will facilitate the acquisition of an L2 lemma through interactive activation: not only will the differences between languages will be gradually acquired, but also the similarities. Translation equivalents in L1 and L2 lead to the co-activation of semantic forms. It can therefore be assumed that if translation equivalence is combined with formal similarity, i.e. if translation equivalents are cognates, the equivalence will be noticed sooner. In the study described below, the formal similarity between lexical items in the L1 and the L2 is one of the main variables included.

The other variable a factor that is independent of the mother tongue and expresses the extent to which the L2 lemma is used: frequency. The role of *frequency* in activation models is obvious, as processing in the lexicon is driven by frequency-induced activation; all elements in the lexicon can attain variable degrees of activation, which increases each time a node is used, and decreases over time. Activated nodes spread activation to nodes with which they are connected. Frequency thus is the major drive behind lexical acquisition.

The two factors described here, input frequency representing the independent effect of L2 on the acquisition process and formal similarity representing cross-linguistic influence, were also the main variables in a recent study on the acquisition of derivational morphol-

ogy in the bilingual mental lexicon (Lowie, 2000). The L2 factor in this study was productivity, which is strongly related to frequency. The study showed that translation equivalence plays a major role in written production at all levels of acquisition, indicating that the learners in this experiment rather strongly relied on their morphological experience in L1. This study also showed that learners, especially at lower levels of proficiency, have not (yet) acquired the productivity of L2 morphological types. Only at the highest level of proficiency included did productivity positively contribute to the scores in the test, and then only in the context where L1 was not explicitly activated. This finding is in agreement with the lexical development that was expected: at early stages of L2 acquisition, the learner's main source of information is his or her mother tongue. At these stages, a full conceptual overlap is assumed between lexical entries in L1 and L2. At later stages, after prolonged exposure to the second language, the restructuring of the semantic form, as exemplified in figure 4, will take place.

Prepositions in English and Dutch

To answer our main research question —to what extent is the order of acquisition related to input (relative frequency of occurrence) or language transfer—we set up an experiment in which the comparative effect of the L1-related variable 'similarity' and the L2-related variable 'frequency' was determined for Dutch learners of English at three different levels of language proficiency.

To keep variables to a minimum, we limited the word classes in the experiment to prepositions only. The choice of prepositions was based on the fact that they occur relatively frequently, so even beginners are familiar with some of them, and that they are easily controlled for meaning in context, so they can easily be elicited. To avoid any problems with figurative or idiomatic uses, the prepositions were used in their most literal, concrete senses, referring to place (e.g. he is *in* his room), time (e.g. he has been ill *for* three months), direction (the ashtray feel *off* the table) possession (the legs *of* the table) and beneficiary (I bought the present *for* my friend) (see Appendix 1 for a full list of the test items used).

A relative frequency list of prepositions was obtained through CELEX (Cobuild Corpus). All occurrences of the prepositions disregarding different senses were taken into account. Prepositions for the test were selected according to the relative frequency, those that occur highly frequently (75.000 or more occurrences in the CELEX/COBUILD corpus, which contains 18,636,970 English lemmas) and relatively infrequently (20.000 or less in the CELEX/COBUILD corpus). From both lists, items were selected that were highly similar in orthography and meaning to their Dutch counterpart when used in a literal sense and those that were not, resulting in the list of items in Table 1. These prepositions were elicited in a cloze test consisting of 25 rather simple English sentences, with the blank to be filled with a targeted preposition (see Appendix 1). To avoid positively affecting "transfer", we did not provide the Dutch equivalent. The test, which took about 15 minutes to complete, was administered during regular class times. All prepositions that fit the English context were considered correct, even if this was not the targeted preposition.

Table 1 Overview of the English prepositions and their Dutch translation equivalents in the four categories selected

	High frequency in English / Dutch equivalent	Low frequency in English / Dutch Equivalent
High similarity to Dutch equivalent	BY / BIJ IN / IN FOR / VOOR ON / AAN	SINCE / SINDS UNDER / ONDER ABOVE / BOVEN
High similarity to Dutch equivalent	AT / AAN TO / AAN BY / DOOR FOR / GEDURENDE TO / NAAR AT / SINDS FOR / VAN OF / VAN FROM	NEAR / BIJ OVER / BOVEN BELOW / ONDER AMONG / ONDER AS FAR AS / TOT BETWEEN/ TUSSEN AMONG / TUSSEN OFF / VAN IN FRONT OF / VOOR

68 Dutch learners of English participated in this experiment. These participants were taken from three naturally occurring groups at the beginning level (a 3VWO class--third year in college preparatory school), the intermediate level (first year non-English majors at the University of Groningen) and the advanced level (third year English majors at the University of Groningen).

Based on previous research and on the literature described above, our expectations were that especially at the lower levels of acquisition a high level of similarity between the Dutch and English prepositions would positively affect the correctness of the answers. The effect of frequency, which was previously shown to occur only after prolonged exposure to the second language, was expected to be strongest at higher levels of acquisition.

Results

The scores of the three groups in this experiment were analysed using a MANOVA, with group as the between-subjects variable (three levels) and similarity and frequency as within-subject variables (two levels each).

The main effect of the between subjects variable, *group*, was significant ($F[2,68]=26.3$; $p<0.01$); the highest number of correct scores was found in the group with the highest proficiency (see Figure 5) Also the main effects of the within-subjects variables similarity $F[1,68]=76.1$; $p<0.01$) and frequency ($F[1,68]=75.2$; $p<0.01$) were both significant.
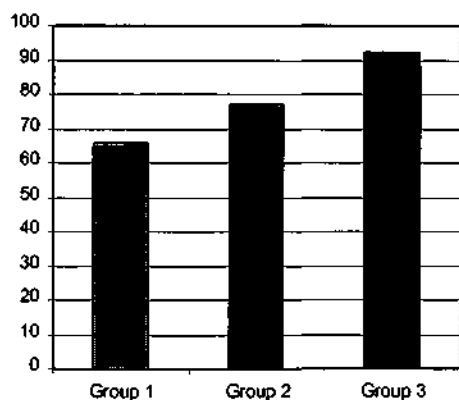

Figure 5 Mean percentage of overall correct scores per group

The interactions between group and frequency (F[2,68]=9.3; p<0.01) was significant (see Figure 6), where the smallest frequency effect was found at the highest level of proficiency.
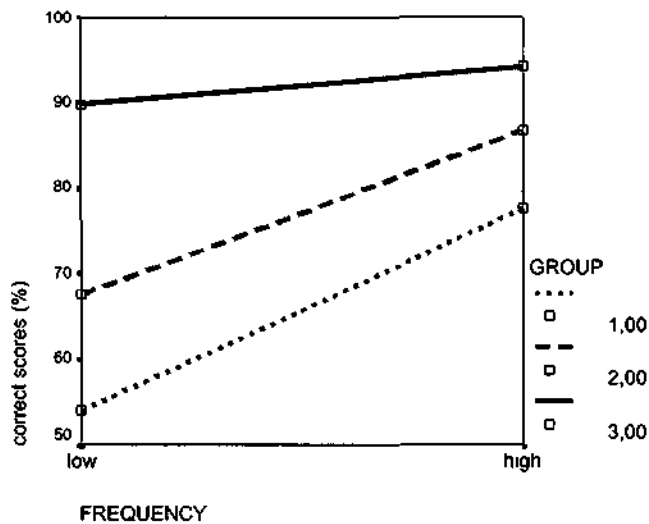

Figure 6 Frequency effect for the three groups in the experiment

Also the interaction between group and similarity turned out to be significant ([F2,68]=10.6; p<0.01). The analysis showed that also for similarity the smallest effect was found at the highest level of proficiency (see Figure 7).

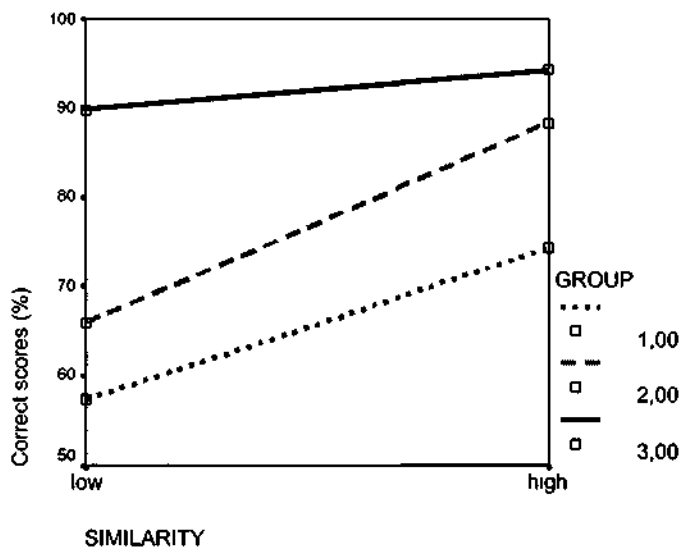*Figure 7 Effect of similarity for the three groups in the experiment*

The interaction between frequency and similarity (F[1,68]=38.8; p<0.01) showed that the largest similarity effect was found with the low-frequency prepositions (see Figure 8).
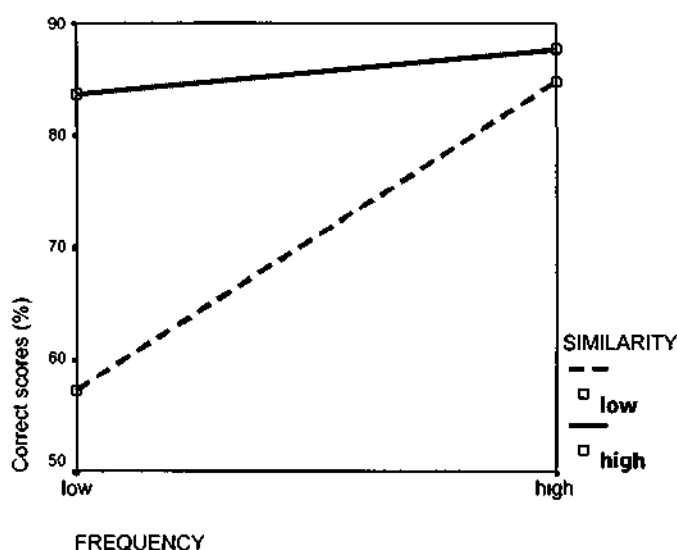
Figure 8 The interaction between frequency and similarity

The three-way interaction between group, similarity and frequency was significant as well (F[2,68]=13.9; p<0.01). The analysis showed that the interaction between similarity (as represented in Figure 7) was only significant for the low and the intermediate level of proficiency, but not for the highest level.

A qualitative items analysis revealed that the lowest overall scores were found with English prepositions that indicate a conceptual distinction that does not occur in Dutch. Examples are the English prepositions *over* and *above*, which are both represented by *boven* in Dutch, and *among* and *between*, which are both represented by *tussen* in Dutch. The same effect, but less strong, was found for *from* and *of*, which are both represented by *van* in Dutch. The number of correct scores for these words was particularly low at the lower levels of English proficiency. The learners typically opted for the more frequent alternative in these cases. Another striking result was that some items that were very similar in both form and meaning in English and Dutch (e.g. *by* – *bij*, as in 'sit by the fire') had comparatively low scores at the lower proficiency levels.

Table 2 Representation of correct scores (%) for some individual items in the three groups

Group	boven / over	tussen / among	van / of	bij / by
1	30	10	40	50
2	12	4	65	65
3	80	29	80	86

Discussion

The experiment showed a clear effect for both similarity and frequency for the low and intermediate levels of proficiency, but hardly any effect for the highest level of proficiency. Considering the high scores of the high proficiency group (see Figure 5), this observation could be attributed to a ceiling effect for the highest level of proficiency. Apparently the subjects in this group had acquired full lexical representations for all the prepositions in the experiment, regardless of degree of similarity to Dutch prepositions or the frequency in the input.

At beginning and intermediate stages, both frequency and similarity tend to affect the score. The effect of similarity was in agreement with our expectations and corroborates results from previous studies. Apparently, the formal similarity to L1 prepositions facilitates the acquisition and use of prepositions in the L2. The finding that also frequency affected the scores for these learners was not in agreement with the expectations. A possible explanation for this finding is that Dutch children are extensively exposed to English. Contrary to the morphological types investigated in previous studies, prepositions are rather frequent. Apparently, the subjects at the lower and intermediate levels of proficiency are sufficiently exposed to the frequently occurring prepositions to show a difference in scores related to frequency levels.

The interaction between frequency and similarity showed that the degree to which L2 prepositions are similar to prepositions in L1 only affected the scores if these prepositions were not very frequent: for the frequently occurring prepositions no effect of similarity was found. The explanation for this finding would be that only for the more unfamiliar prepositions the subjects tend to rely on their first language. This explanation is supported by the observation that this interaction did not occur at the highest proficiency level: these students had been sufficiently exposed to the L2 to have developed full representations for all prepositions.

The results of the qualitative data analysis clearly seem to indicate the developmental process predicted by the model (and exemplified in Figure 4). At the lower proficiency levels most subjects show a presumed conceptual overlap between the L1 and the L2 lexical items. At the highest proficiency level almost all subjects seem to have acquired the additional conceptual category in English. Only the difference between the prepositions *among* and *between* has not (yet) been acquired by the majority of the subjects at the highest proficiency level. The striking finding that in some cases low scores were found while both formal similarity and conceptual overlap were high seems to point to the learners' reluctance to use formally similar words for the translation of metaphorically used lexical items (cf. Kellerman 1987).

Conclusion

In this paper we have summarised an interactive activation model of the bilingual mental lexicon and presented an empirical study that investigates one aspect of this model. The model summarised here claims that all overlap between items in the mental lexicon is mediated by the conceptual representations. Different language subsets can be assumed by sets of lexical items that share the same language property. Through a mechanism of interactive activation, both lexical subsets and individual lexical items can have different levels of resting activation. The process of acquisition of lexical items is a matter of gradual development from assumed full conceptual overlap between L1 and L2 items at beginning stages of acquisition to native-like L2 representations at advanced levels of acquisition.

The main purpose of this study was to determine the role of the learner's first language in the acquisition of lexical items in a second language and to investigate how this role develops over time. We argued that the first language plays a role at two levels of the model described: at the level of semantic overlap between lexical items in the two languages, mediated by the conceptual representations, and at the formal (phonological / orthographic) level. The focus of the current paper was on the latter type of cross-linguistic influence. The amount of formal overlap between the two languages was represented as the different levels of formal similarity between Dutch and English prepositions. The experiment shows that formal similarity plays a role at the beginning and at the intermediate level of proficiency, but not at the highest level of proficiency. This finding is in agreement

with what we expected to find. We hypothesised that beginning learners will presume a full overlap between lexical items in the two languages and that the learner's assumptions are reinforced by formal similarity of L1 and L2 lexical items. This shows that at early stages of L2 acquisition learners rely on their first language in learning and using L2 lexical items. Only at the highest level of acquisition has the learner developed complete L2 entries and does no longer have to rely on their first language.

On the basis of previous research on the acquisition of L2 derivational morphology, we also expected that frequency would only play a role at the highest levels of L2 acquisition. The reason for this was that the effect of frequency was expected to start affecting L2 performance only after prolonged exposure to the second language. The experiment, however, showed that frequency did play a role at the lower levels of L2 acquisition, but not at the highest levels. We accounted for this observation by arguing that different from morphological types in earlier experiments, the average frequency of prepositions is comparatively high. This results in high frequency prepositions that are abundant even in the beginning learner's input. The absence of an effect for the highest level of acquisition could be explained by a ceiling effect: the scores for this group on prepositions was already very high for the low frequency prepositions, leaving no room for a further frequency effect.

Although the findings in the study reported here are clearly significant, the results must be interpreted with care. The subjects in this study were taken from three intact classes and their number was limited. However, for the main effects of significance and frequency, this is no serious drawback, as these were within-subjects variables. And although the cross-sectional design with these groups cannot be considered identical to real development as measured in a longitudinal design, the clear difference between the cross-sections taken gives a strong indication in the direction of the development observed.

This study strongly suggests that up to the intermediate level of acquisition learners strongly rely on formal similarities with their L1 in using L2 prepositions. Further research will have to determine whether the translation equivalence of lexical items, i.e. the amount of conceptual overlap between L1 and L2 lexical items, is noticed if it does not coincide with formal similarity. It can be assumed that the greatest difficulty for learners can be found in cases where formal similarity does not coincide with translation equivalence. This category (of "false friends") needs further investigation with learners at different stages of L2 acquisition. Another interesting observation that merits further investigation is that an items analysis of the current experiment revealed that English prepositions that indicate a distinction that does not occur in Dutch get the lowest scores (e.g. *tussen = between / among*; *boven = over / above*). The difference between the groups for the scores on these items seems to support the gradual restructuring of L2 lemmas from assumed conceptual overlap with similar L1 items to independent (though partially overlapping) L2 lemmas.

The experiment described here is by no means sufficient to support (or falsify) the entire model of the bilingual mental lexicon described here. Only one aspect of the model has been tested and many other questions are yet to be answered. However, with this experiment we hope to have contributed to the gradually emerging picture of the bilingual mental lexicon that can only be established by the constant interaction between empirical studies and theoretical modelling.

References
Clark, E.V. (1993). *The lexicon in acquisition*. Cambridge: CUP.
Jang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics 21* (1), 47-77.

Kellerman, E. (1987). *Aspects of transferability in second language acquisition.* Unpublished Ph.D. thesis, University of Nijmegen.

Kroll, J.F. (1993). Assessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 52-81). Amsterdam/Philadelphia: Benjamins.

Levelt, W.J.M. (1989). *Speaking: from intention to articulation.* Cambridge Mass./London: Bradford.

Lowie, W.M. (1991). Derivationele morfologie en tweede-taalverwerving. *Toegepaste Taalwetenschap in Artikelen* 39, 83-92.

Lowie, W.M. (1998). *AIM: The Acquisition of Interlanguage Morphology.* Ph.D. thesis, Rijksuniversiteit Groningen.

Lowie, W.M. (2000). Cross-linguistic influence on morphology in the bilingual mental lexicon. In *Studia Linguistica 54, 2,* 175-185.

Schreuder, R. en Baayen, H. (1995). Modelling morphological processing. *Morphological aspects of language processing.* ed. Feldman, L.B., 131-156. Hillsdale: Lawrence Erlbaum.

Woutersen, M. (1997). *Bilingual word perception.* Ph.D. Thesis, University of Nijmegen. Nijmegen: Katholieke Universiteit Nijmegen.

Appendix 1 – full representation of test items used in the experiment

The quiz contained a blank for the underlined preposition. The Dutch equivalent was <u>not</u> provided (but is given here to show level of similarity).

1. Let's sit <u>by</u> (BIJ) the fire.

2. He is <u>in</u> (IN) her room, I suppose.

3. I bought this present <u>for</u> (VOOR) my friend.

4. There is a picture <u>on</u> (AAN) the wall.

5. I have liver here <u>since</u> (SINDS) 1995.

6. The cat is lying <u>under</u> (ONDER) the table.

7. The plane was flying <u>above</u> (BOVEN) the clouds.

8. He is <u>at</u> (AAN het) work

9. I will never lend my car <u>to</u> (AAN) my son.

10. She was killed <u>by</u> (DOOR) a bullet.

11. He has been ill <u>for</u> (GEDURENDE) three months.

12. He went <u>to</u> (NAAR) London for a week.

13. They threw tomatoes <u>at</u> (NAAR) him.

14. He has been absent <u>for</u> (SINDS) three days.

15. The legs <u>of</u> (VAN) the table are wobbly.

16. I knew I could trust him <u>from</u> (VAN) the moment I saw him.

17. He lives <u>near</u> (BIJ) Cambridge

18. He was leaning <u>over</u> (BOVEN) the body when he was struck on the head.

19. His apartment is <u>below</u> (ONDER) mine.

20. <u>Among</u> (TUSSEN) my friends, there is no one who can help me.

21. To get to the station, walk <u>as far as</u> (TOT) the traffic lights, and then turn right.

22. His house stands <u>between</u> (TUSSEN) his sister's and mine.

23. I couldn't find my friend <u>among</u> (TUSSEN) all those students.

24. The ashtray fell <u>off</u> (VAN) the table.

25. I saw somebody standing <u>in front of</u> (VOOR) your window, trying to look inside.