The acquisition of interlanguage morphology

Lowie, Wander Marius

Chapter 4

An empirical investigation

4.1 Introduction

This chapter investigates some of the predictions made on the basis of the theoretical observations in the previous chapters. Not all aspects of the model are equally suitable to be empirically tested or can be tested within the scope of the current study. Therefore, some aspects have been selected for investigation, all focused on determining the factors that affect the psychotransparency of morphologically complex words in a second language. First, an exploratory study is reported on (4.2) that compares the impact of form-based similarity of L1 affix types and L2 affix types to the impact of semantic similarity between these types. Starting from the results of this exploration, two experiments were conducted that investigate to what extent syntactic and semantic overlap of L1-L2 affix pairs affect psychotransparency. In one of these studies, it was attempted to test the effect of interlingual activation in a priming experiment involving reaction time measurement (4.3). The other studies contain two steps. First, the degree of translation equivalence between L1 and L2 affix types was determined in a corpus study. Second, the outcomes of the corpus study were tested in an experimental setting involving the effect of different levels of translation equivalence and productivity in an L2 production task (4.4). The overall conclusion of these studies is provided in 4.5. The languages of investigation in all studies described here are L1 Dutch and L2 English.

4.2 An exploratory study of L2 morphology

4.2.1 Introduction

In Chapter 3 it was claimed that psychotransparency (including word-internal and individual factors) is an essential condition for the analysis from item to type. This claim is supported by studies into L1 acquisition, as described in 3.2. One of the conclusions drawn by Clark (1993), for instance, was that of the strategies children use in creating new words, transparency takes precedence over productivity and frequency, and that transparency is one of the main factors determining the acquisition of lexical items.

In 3.4.3 it has been argued that the main difference between L1 learning and L2 learning is that L1 learners will gradually acquire the conceptual characteristics as-

sociated with the semantic form of word formation types as well as their corresponding lexemes. L2 learners, on the other hand, have already acquired the concepts, and will "only" have to relate these to the correct types and corresponding L2 lemmas. As was argued there, this means that there are three areas of problems the L2 learner has to cope with. Firstly, learners may have problems with L2 types for which no equivalent L1 form exists and, vice versa, with L1 types for which no corresponding L2 form is available. An example is the Dutch affix *-sel* as in *zaagsel* (sawdust):

(7) [[V_{dyn}]____] [N, -abstract] ('that what remains after Ving').

No equivalent English affix form represents this type. Secondly, the problem of polysemy and synonymy is multiplied for L2 learners: there may be L2 forms that are similar to L1 forms, but that do not represent the same type, and different types may be represented by similar forms. For Dutch and English this was exemplified by the affix *-ster*: in English, this form refers to an agent, male or female ('person of a certain type or of a certain trade or interest'), while in Dutch it refers to female agents only. Finally, it will be very hard for L2 learners to acquire the subtle differences that distinguish certain (similar, but not identical) L2 types. The example given in 3.4.3 was the difference between the two types represented by *-ful*: the acquisition of the subtle differences between the types in L2 is further complicated by the lack of a consistent translation equivalent for these types. Moreover, several studies concerning the acquisition of L2 morphology have shown that learners have most difficulty in dealing with "deceptive transparency": words that seem transparent, but are opaque in fact.

Taking psychotransparency as a starting point, four categories of affix types can be distinguished for L2 learners:

A. Semantically equivalent types that are represented by identical or very similar forms in L1 and L2. Words based on these types can be expected to be transparent for L2 learners, independent of the learners' knowledge of L2 morphology, since the types fully overlap in terms of form and semantic content. An example for Dutch learners of English is agentive *-er* (lezer - reader).

B. Semantically equivalent types that take a different shape in L1 and L2. The transparency of words based on these affix types is dependent on the learner's familiarity with the L2 form. Once the learner has made a link between the two types, the semantic overlap between the L1 and L2 types will facilitate the use of forms based on these types, especially if there is a consistent translation equivalent for in the affix in L1. Example: Dutch *-heid* - English *-ness* (open*heid*, open*ness*).

C. Similar forms in L1 and L2, which are based on semantically different types. These forms can be labelled "deceptively transparent": they seem to be identical, but are not. An example for Dutch learners of English is the affix *-ster*, which in Dutch always denotes a female agent, but can be either male or female in English: Dutch *omroepster* -a female announcer-, vs. English *speedster* -male or female-. Other affix types that fall within this category are L2 types that lack a consistent translation equivalent in the L1 and for which a similar Dutch affix

leads to deceptive transparency. An example is the English affix *-ful*, which in some words is the translation equivalent of the similar Dutch form *-vol*, but in many other words is represented by different affixes in Dutch.

D. L1 types for which no equivalent L2 forms exists or vice versa. The transparency of words in this category is completely dependent on the learner's familiarity with the L2 type and form. An example of this type would be Dutch *-sel*, worked out in (1) above.

The acquisition of affixes of the first three types was investigated in an exploratory study of L2 morphology[42], consisting of three experiments. The main purpose of this study was to serve as a pilot study on which further research could be built. After a brief outline of the method of investigation (4.2.2), the results will be represented of each of the three sub-tests of this pilot study, followed by a brief discussion of each of the sub-tests (4.2.3 to 4.2.5). Then the results of all three sub-tests will be reflected upon in a more elaborate global discussion of all the results (4.2.6), followed by a general conclusion of this study (4.2.7).

4.2.2 Method

From each of the categories above, three suffixes were selected (see Table 4) and presented to 34 Dutch secondary school pupils from two different levels of L2 acquisition (third form and fifth form of the "VWO", pre-university education), aged approximately 14 and 16 respectively (see details in 4.2.2.1). Three sub-tests were administered. Instructions and excerpts from all tests have been included in Appendix 1.

The first test was a two-part translation test in which learners were asked to provide English equivalents for morphologically complex Dutch words (e.g. doofheid -), and Dutch equivalents for morphologically complex English words (e.g. deafness -). To minimise effects of item-familiarity, the translation of the stem was given (e.g.: doof = deaf). 36 items were divided over the two tasks, eight representing category A, eleven for B, sixteen for C. Eleven morphologically complex decoys were added to the scoring forms, which consisted of transparent prefixed words and opaque suffixed words; these decoys were not included in the analysis.

The second test was a judgement task, in which learners had to decide whether the items in the test were valid English words. In the instructions, the subjects were told that the words in the list had to be marked "possible" or "not possible". Furthermore, it was emphasised that we were not after "existing" words, but after (morphologically) "possible" words. This was illustrated with some examples. This test consisted of 26 items; nine from category A, six from B and nine from C. Six decoys were added to the scoring forms, which were not included in the analyses. Half of

[42] Category D was not included in the experiment, because two problems for the analysis were anticipated. First, this category would have to depend on a very limited number of affixes. Second, the interpretation of the scores would be problematic, as it is not clear what should be regarded as a "correct" score.

the items consisted of "possible" words, the other half of "impossible" words. All "impossible" words were pseudo-words that were illegal due to the incorrect or awkward affixation. For instance, affixes were attached to stems belonging to syntactic categories or subcategories that were not in agreement with the subcategorisation restrictions of the affix type (e.g. *softity*). The words in Category C were subject to deceptive transparency based on the learners' L1 (e.g. *learnsome*). To enable controlling for individual variation, a Dutch control test was included, in which learners had to decide on the possibility of a set of morphologically complex Dutch words and pseudo words. The Dutch sub-test was administered before the English sub-test.

The third test was an analysis test in which learners had to select the best possible meaning for a given pseudo-word in a multiple choice test ("frickless = looking like a frick/without frick/something that can be fricked"). Besides, the learners were asked to select a syntactic category of the each pseudo-word. This test contained four affixes from categories A to C. The focus of attention was rather obviously on the meaning, function and syntactic category of the affix types. One of the alternatives in the semantic multiple-choice answers related to all items was "not possible in English". The syntactic categories the learner could choose from were Verb, Noun and Adjective. These word class labels were provided in Dutch. Similar to Experiment 2, this experiment consisted of an English sub-test and a Dutch sub-test to enable controlling for individual differences. The Dutch sub-test was administered before the English sub-test.

Instructions for all tests only concerned the way in which the forms had to be filled in and no additional information was provided about the actual issue of investigation. The order in which these test were presented ranged from little attention to morphological complexity in the translation test to abstract, very explicit questions about the morphological structure of words and the functions and meanings of morphemes in the final test (see discussion in 3.2.2).

Table 4. Affixes selected for each category in the test.

CAT A		**CAT B**		**CAT C**	
Dutch	*English*	*Dutch*	*English*	*Dutch*	*English*
-er	-er	-baar	-able	-dom	-dom
-iteit	-ity	-heid	-ness	-ster	-ster
-loos	-less	-(acht)ig	-ish	-ful	-vol

4.2.2.1 Subject groups

The two subgroups in this experiment are assumed to represent different stages in the acquisition of L2 English. The learners from the fifth form can be expected to have reached a higher level of English proficiency than the learners from the third form, as they will have had more years of English instruction and have been exposed to English more. However, these learners do not only differ in level of L2 proficiency, but also in age. To determine the relative effect of these factors, a questionnaire was administered in which enquiries were made about the subjects' age and years of English instruction. The subject groups differed significantly on age

(χ^2=28.9; *df*=5; p=0.00002) and on years of English Instruction (χ^2=18.8; *df*=7; p=0.009). The distribution of these variables over the groups is graphically represented in Figure 21. Interviews with the English teachers of these groups revealed that during classes hardly any attention had been paid to morphological generalisations.

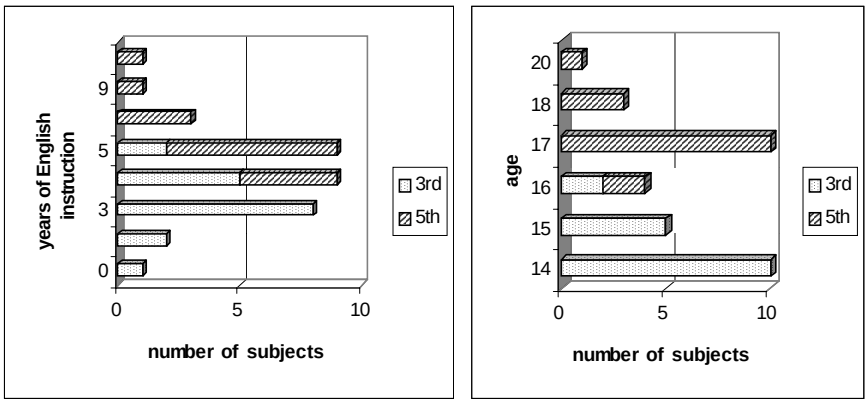

Figure 21. The distribution of two between-subject variables over the two subjects groups in these experiments.

An additional between-subject variable in these experiments was Exposure to English. This was determined by a set of questions about experience in English outside school, like English relatives, holidays in English speaking countries and the approximate exposure to English spoken media. The distribution of this variable across the subject groups is represented in Figure 22. The two subgroups did not differ significantly with regard to the exposure index (χ^2 =6.6, *df*=5; p=0.26).

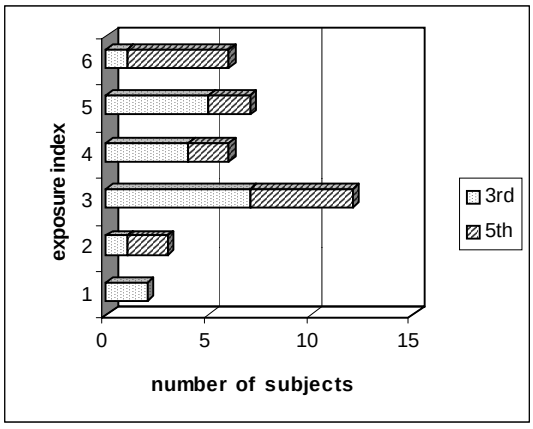

Figure 22. The distribution of the six levels of exposure index over the two subgroups.

4.2.3 Experiment 1

4.2.3.1 Introduction

The translation experiment was analysed on three variables: the differences between the two subject groups, the categories (A to C) and the direction of the translation (Dutch to English and English to Dutch). From an L2 perspective, the direction of translation can be seen as a production task and a comprehension task respectively, and are referred to as such in this section. In addition to these analyses, the interaction was investigated between the categories and the subgroups. A motivation for these analyses is given in the sections concerned.

The model outlined in the previous chapters would predict that the sequence of acquisition of morphological types is affected by the inherent transparency of the morphological types, the psychotransparency as perceived by the learner, productivity, frequency and simplicity. An increase of exposure to English implies a greater chance of being familiar with morphological types. The same will apply to a difference in years of instruction in English: the more instruction in English, the greater the chance of being familiar with morphological type, even though no explicit attention had been paid to morphological generalisations in instruction. Therefore, the general knowledge of morphological types can be expected to be more extensive for the learners in the fifth form. The age factor included in this experiment would imply a difference in L1 maturation, which can be expected to increase the possibility of transferring morphological generalisations (see 3.4.4).

In this experiment, the learner's native language will particularly play a role at the level of the psychotransparency of morphological types: the affix types that display a greater form-based similarity to a type in the learner's L1 can be expected to be more transparent. In terms of the model outlined in the previous chapter, form-based similarity will lead to co-activation at the level of lexemes (see 3.3.2), which yields a facilitating effect if it coincides with semantic overlap. Therefore, the affix types in Category A were expected to be acquired and applied more easily than the ones in Category B. The deceptively transparent types in Category C were expected to be most difficult to apply. Due to the form-based similarity combined with the lack of consistent translation equivalents for these types, the error rate in category C can be expected to be relatively high.

The difference between the two tasks in this test (comprehension vs. production) can be expected to yield a difference in scores. In the previous chapters (see 2.5.4 and 3.2.3.2), it has been argued that, for production, fully specified lexical entries are required, while comprehension can occur based on incomplete lexical entries. Applied to the current experiment, it can be expected that a learner is able to recognise a particular L2 word, but is not (yet) able to produce it due to under-specified lemma nodes. Therefore, the overall score for comprehension is expected to exceed the overall score for production.

An interaction may be expected between the amount of exposure to English and the interpretation and production of morphologically complex words, as comprehension can be expected to precede production. At higher levels of exposure, the differ-

ence between comprehension and production can be expected to be smaller than at lower levels of exposure. Furthermore, an interaction can be expected between the amount of exposure to English and the affix categories in this test, as facilitation due to L1 similarity will affect the order of acquisition of the affix types in the categories (A before B before C). Consequently, the difference between the categories will increase with increasing exposure.

4.2.3.2 Scoring

To all of the subjects' scores in Experiment 1 a response-code was assigned, comprising the following information:

- Correctness of the affix used, based on the Random House Webster's Unabridged Electronic Dictionary.
  Example: *fathership* (for Dutch "vaderschap") was regarded as incorrect; *fatherhood* as correct. Non-morphologically complex forms not regarded as "correct" responses, as they did not comply with the assignment.
- Correctness of the syntactic category of the affix type.
  Example: the syntactic category of *reachsome* for *bereikbaar* ("reachable") was considered correct; that of *reachment* was not.

For incorrect responses the following information was added:

- The origin of the error (L1 or L2).
  Example: *valueful* was considered as based on L1 (Dutch "waardevol"), *besiegance* (for *besiegement*) was considered as based on L2, as Dutch only has *omsingeling*.
- Meaning-based or form-based error.
  Example: *girly* for *meisjesachtig* ("girlish") was considered meaning-based as the affix bears no form-based similarity to *-achtig* or *ish*; *valueful* was considered to be form-based, due to its obvious orthographic and phonological similarity to the Dutch affix *-vol*.

4.2.3.3 Results

Two analyses were used to investigate the data scored as described above. Firstly, a MANOVA test was run to determine the relative effect of the variables involved on the percentage of correct scores. In this analysis, Category (A-C) and Task (Production-Comprehension) were used as within-subject factors. The mean scores for each group is represented graphically in Figure 23. Secondly, several χ^2 analyses were run to test the differences between the frequencies of correct, incorrect and blank scores for the different variables involved.

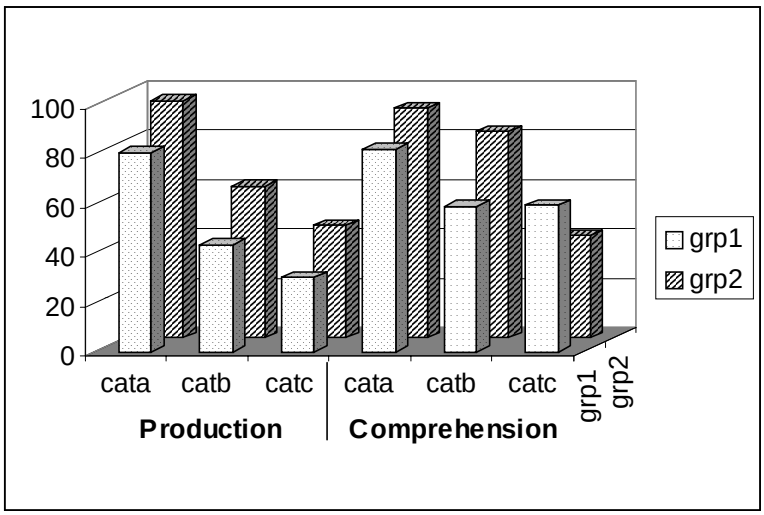

Figure 23. Representation of the mean percentage of correct scores in Experiment 1. The scores for each category are separately represented for the two tasks: production and comprehension.

4.2.3.3.1 Between-subject effects

The overall difference between the subject groups turned out to be significant at $p<0.01$ ($F[1,32] = 11.32$; $p=0.002$). Further analyses were applied including the number of years of instruction in English ($F[6,26]=1.98$; $p=.106$), and age ($F[5,27]=7.28$; $p=.011$). Finally, a design was run testing the exposure index (as defined by the factors mentioned above, which turned out not to be significant at $p<0.05$ ($F[3,30]=1.35$; $p=.277$). The difference between the subgroups in the type of answer (correct, incorrect and blank) is shown in Table 5. In this table, no additional subdivision was made for meaning-based and form-based errors; errors based on L1 were mostly form-based; errors based on L2 were mostly meaning-based. Answers that avoid the target word by description are clustered together under "non-morphologically complex" at the bottom of this table. The differences turned out to be significant at $p<0.01$ ($\chi^2=88.9$; $df=12$).

Table 5. Frequency table of the differences between groups. In parentheses is the percentage of all scores for each sub-group (column percentages).

	3rd form (n=629)	5th form (n=576)
Blanks	108(17.6)	16 (2.9)
Correct response	321(52.5)	389(69.5)
Incorr, correct syncat.	134(21.9)	126(22.5)
Of which L1 based	107 (17.5)	101 (18.0)
Of which L2 based	27 (4.4)	25 (4.5)
Incorr, incorr syncat	15 (2.5)	16 (2.9)
Of which L1 based	7 (1.1)	4 (0.7)
Of which L2 based	4 (0.7)	8 (1.4)
Of which ambiguous	3 (0.5)	3 (0.5)
Non-morph complex	34 (5.6)	13 (2.3)

4.2.3.3.2 Categories

The main issue of investigation in this pilot was the difference between the affix categories (see 4.2.1) A to C as within-subjects factor. The difference between the categories (see Figure 23) was significant at $p<0.01$ ($F[2,64]=114.67$; $p=0.000$). The differences between the subgroups in terms of correct, blank and incorrect scores, summarised in Figure 24, Figure 25 and Figure 26, were found to be significant in all categories (χ^2 = 9.6, 28.7 and 41.4 for Category A, B and C respectively; $p<0.01$ in all cases).

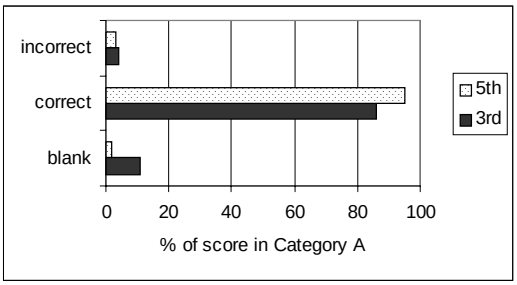

Figure 24. Percentage of incorrect, correct and blank scores in Category A across tasks.

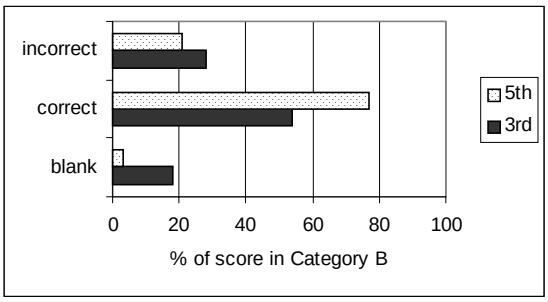

Figure 25. Percentage of incorrect, correct and blank scores in category B across tasks.

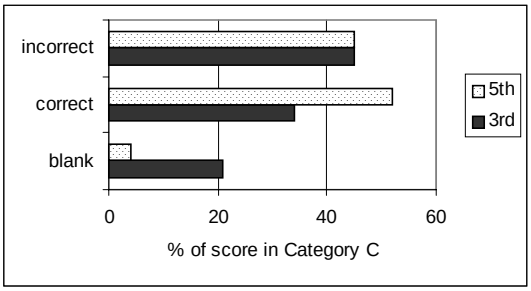

Figure 26. Percentage of incorrect, correct and blank scores in category C across tasks.

4.2.3.3.3 Task (production/comprehension)

The difference between production and comprehension was analysed to investigate a possible asymmetry between these tasks. The overall difference between the sub-tests, as shown in Figure 23, turned out to be significant at $p<0.01$ ($F[1,32]=12.41$; $p=0.001$). The difference between the subgroups in terms of blank, incorrect and correct scores was significant in both tasks $\chi^2 = 39.1$ and 40.6 for production and comprehension respectively; $p<0.01$ in both cases. These data are represented in Figure 27 and Figure 28.

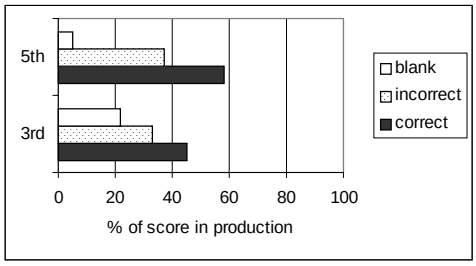

Figure 27. Percentage of blank, incorrect and correct scores in production task across categories.

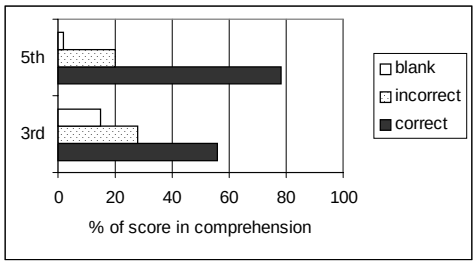

Figure 28. Percentage of blank, incorrect and correct scores in comprehension task across categories.

4.2.3.3.4 Interactions

A significant interaction was found for Task x Category ($F[2,64]=4.94$; $p=0.01$), where the difference between scores on production and comprehension was largest in Category C. No significant interactions were found for any of the between-subject variables included in this experiment.

4.2.3.4 Discussion

The difference between the two subject groups, representing different levels of L2 acquisition, turned out to be significant. An interesting question, however, is what determines the difference between these two naturally occurring groups. An explanation might be found in some of the between-subject factors included in the design. Since the subgroups differed significantly in terms of age and years of English instruction (see 4.2.2.1), the explanation was to be found in a difference regarding these variables. The factor age turned out to be significant; older learners obtained higher scores. Since the amount of formal instruction in English did not turn out to be a significant factor, another age-related factor must be involved. This seems to imply that command of L1 morphology is a more important factor in the acquisition and use of L2 morphological types than either instruction in the foreign language or

exposure to the foreign language. In the χ^2 analysis, the largest differences observed between subjects were a generally larger number of correct responses at the higher levels of acquisition and the larger number of blanks at the lower level of acquisition. Furthermore, the subjects from the third form used more descriptions than did the subjects from the fifth form. These observations can be explained in terms of the development of morphological knowledge in L1; learners with a greater command of L1 morphology have acquired more morphological types and analyse and more often produce morphologically complex words type-familiarly. Learners at the lower level use different strategies to make up for their lack of type-familiar knowledge: they leave out the item altogether (blanks) or they provide descriptions ("non-morphologically complex"). Surprisingly, exposure to English, operationalised as the Exposure Index, did not turn out to be an accurate predictor of L2 morphological performance.

The difference between the categories turned out to be very clear. There were hardly any incorrect responses in category A (about two per cent), there were significantly more errors in category B, while category C showed the largest number of incorrect responses. The fact that the largest number of blank scores was observed for the affix types in Category C confirms the idea that learners experience the types in this category as the most difficult. These observations are in line with the expectations mentioned in section 4.2.3.1.

Contrary to what had been expected, the interaction between the exposure and the percentage correct scores between the categories and between the tasks was not significant. Apparently, the difference in exposure to English between these subgroups was not sufficiently large for this effect to show.

The fact that the total number of correct responses for comprehension was larger than the number of correct responses for the production sub-test confirms the expectations on this variable. More interesting, however, is to look at the difference between the subgroups at this variable. If comprehension preceded production, and if the acquisition of the morphological types concerned were acquired in the time span investigated in this experiment (between the third and the fifth form), an interaction between the tasks could have been expected. In that case, there should be little difference between the subgroups with regard to comprehension and a larger difference with regard to production. The data presented above, however, reveal that this interaction was not found (see, e.g., Figure 23). Apparently, the distance in level of acquisition between these subgroups is not big enough to show the difference. An alternative explanation for the lack of interaction might be that the difference is lost in a stronger facilitating effect of the comprehension task. Yet, no conclusive answer can be given to this question.

4.2.4 Experiment 2

4.2.4.1 Introduction

The scores in Experiment 2 were encoded for correctness. This experiment consisted of a Dutch part and an English part. The Dutch part was included to investigate the

ability of the individual subjects to interpret morphologically complex words in their mother tongue type-familiarly. As the affix categories were based on the expected difference in psychotransparency due to L1/L2 similarity, no categories could be distinguished in the L1 test. In the English part, the categories as described in Table 4 were distinguished. Analyses were carried out to determine the effect of the subgroups in the two sub-tests. Further factors entered in the analysis were the categories of affix types (A to C), the interaction of the between-subject variables (exposure index, age, years of instruction in English and group) and the categories, and the interaction between the Dutch and the English sub-tests. Similar to Experiment 1, it was to be expected that the scores would differ between the two subgroups in this experiment. For the same reasons as mentioned in the description of Experiment 1, the largest proportion of correct answers in the English sub-test were expected to occur in Category A, followed by B, followed by C. Furthermore, within the categories the largest difference between the subgroups was to be expected within categories B and C, as the developmental aspect is most obvious for these categories. Finally, it was expected that the subjects would score better on the Dutch sub-test than on the English sub-test, as all subjects will have fully acquired most of the morphological types in this experiment in their native language, but not necessarily in their L2.

4.2.4.2 Results

An overview of the results is represented graphically in Figure 29.

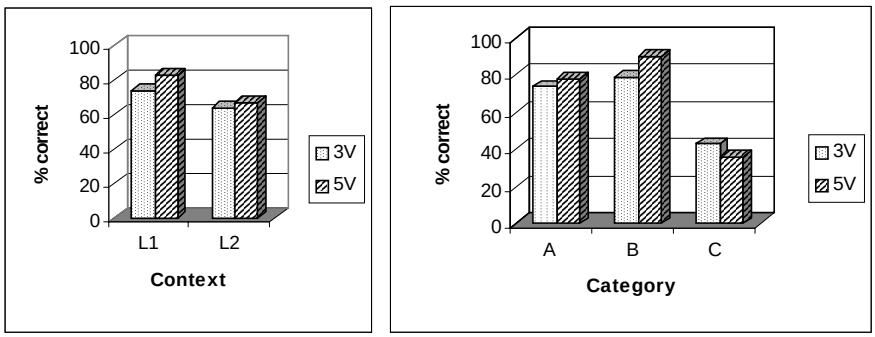

Figure 29 Graphic representation of the mean percentage of correct scores in Experiment 2. The figure on the left-hand side indicates the difference between the subjects' scores in the two sub-tests. The figure on the right-hand side represents the difference between the three categories distinguished in the L2 context.

4.2.4.2.1 Between-subject effects

The difference between the two subgroups turned out not to be significant, neither between the sub-tests ($F[1,31]=3.43$; $p=0.073$) nor between the categories ($F[1,231]=0.68$; $p=.42$). Of the other between-subject factors, neither the exposure index nor the age was found significant in this experiment. However, a significant

effect was found for the years of English instruction between L1 and L2 (F[6,25]=3.12; p=0.020.

4.2.4.2.2 Categories

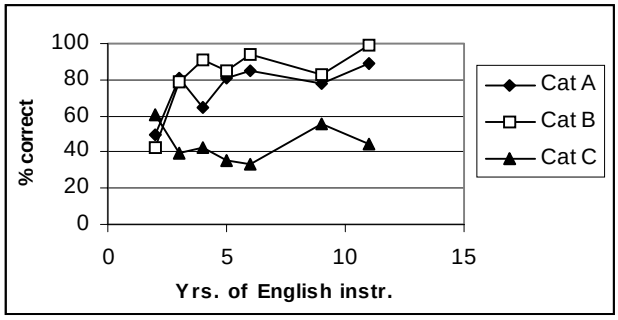

Figure 30. Percentage of correct scores as a function of the years of instruction in English.

The difference between the categories (in the English sub-test) turned out to be significant (F[2,62]=52.93; p<0.001). Moreover, a significant interaction was found for years of English instruction and Category (F[12,50]=2.14; p=0.031), where the effect of years of instruction in English was strongest for the items in Category B, followed by A and C respectively (see Figure 30).

4.2.4.2.3 Dutch and English sub-test

The difference between the English and the Dutch sub-test turned out to be significant (F[1,31]=46.77; p<001). The correlation between the subjects' scores on the L1 part of the test and the L2 part of the test was significant at p<0.01 (r_{xy}=.48).

4.2.4.3 *Discussion*

No overall effect was found for the sub-groups within the English sub-test, nor between the two sub-tests. However, the years of English instruction was a significant factor in the MANOVA involving the scores on the two sub-tests. In itself, this makes sense, as it can be expected that the amount of instruction has an effect on L2 morphological performance. But the fact that no interaction was found between the English and the Dutch sub-test on the one hand and years of instruction on the other implies that performance on Dutch morphology also increases with years of instruction of English. This effect cannot be attributed to a difference in age or exposure to English. From this it may be tentatively concluded that learning a second language affects the skill to apply word formation types in L1, possibly due to raised awareness leading to an increased psychotransparency of L1 words[43].

[43] The influence of L2 learning on L1 awareness has been suggested earlier by, for instance, Vygotsky, who cites Goethe in: "he who knows no foreign language does not truly know his own." (1962:110)

The findings about the difference between the categories were roughly in line with what had been expected: most errors were encountered in Category C, while significantly fewer errors were found in Categories A and B. In this experiment, however, the difference between A and B was minimal. This suggests that the form-based similarity between L1 affix type and the L2 affix type is not relevant in this type of experiment. The interaction between the years of English instruction and the percentage of correct scores on the different categories points out that the development of L2 morphology predominantly takes place for category B and (to a lesser extent) category A rather than category C. This observation is in agreement with the expectations formulated in 4.2.4.1.

Finally, some observations were made about the subjects' performance in English and Dutch morphology. Both subject groups scored better at the Dutch sub-test, which confirms the expectations. Obviously, learners are better at applying morphological types in their L1 than in the L2. The fact that a significant and relatively strong positive correlation was found between the subjects' scores in both tests, indicates that individual differences in the application of morphological types did play a role in this experiment.

4.2.5 Experiment 3

4.2.5.1 Introduction

The variables included in the morphological assessment experiment were, again, the categories of affix types, the Dutch and the English sub-test and the between-subject variables group, age, years of instruction in English and exposure. An additional dependent variable that had been explicitly included in this experiment is the score on syntactic categories.

More than the other experiments, this experiment investigates the learners' metalinguistic awareness of morphology and their ability to reflect on this rather than actual "linguistic" knowledge (see the instructions and the excerpts in Appendix 1). Due to the high level of abstractness of this experiment, it was expected that the difference between the two subject groups would be larger than in the other experiments discussed so far. After all, the subject groups do not only represent the level of L2 learning, but also a difference in age. For the native language, it has been argued in Chapter 3 that "knowledge precedes awareness" (see 3.2.2.2). Therefore, older learners are more likely to have acquired morphological awareness (in L2 and in L1).

The difference between the categories was expected to be similar to the outcome in the previous experiments. The largest proportion of correct scores were expected to occur in Category A, followed by B and C respectively. Between the subgroups, the least difference was expected to occur in category A, as words in this category would be equally transparent for both subgroups.

Differences were expected between the Dutch and the English sub-test, as learners are more likely to have acquired all conceptual implications of the morphological types in their native language than in a foreign language. However, a complicating

factor is that this experiment concerns metalinguistic awareness. Learners may have acquired morphological awareness through internal analysis, but also through formal instruction. As the learners in this experiment had not been given specific training in the analysis and production of morphologically complex words in English, this could not be seen as an advantage over the discovery of Dutch morphological types. The amount of instruction being equal, the learners in this experiment can be expected to have a greater morphological awareness in their native language than in English.

Finally, differences could be expected between the scores on the semantic and the syntactic parts of this experiment. The attribution of syntactic categories to morphologically complex pseudo-words not only requires metalinguistic awareness of the morphological type concerned, but also of the linguistic terminology and the application of that terminology. Therefore, the proportion of correct scores on the syntactic categories can be expected to be generally lower than the number of correct semantic interpretations of the words in the experiment. Moreover, the learners from the fifth form can be expected to be more familiar with the terminology than the learners from the third form, as they will have received more instruction in syntax. In addition, the learners in the fifth form can be expected to be more able to deal with the high level of abstractness for the syntactic part of the experiment.

4.2.5.2 Results

A graphic representation of the mean percentage of correct scores is represented in Figure 31 (semantic scores) and Figure 32 (scores on syntactic categories).

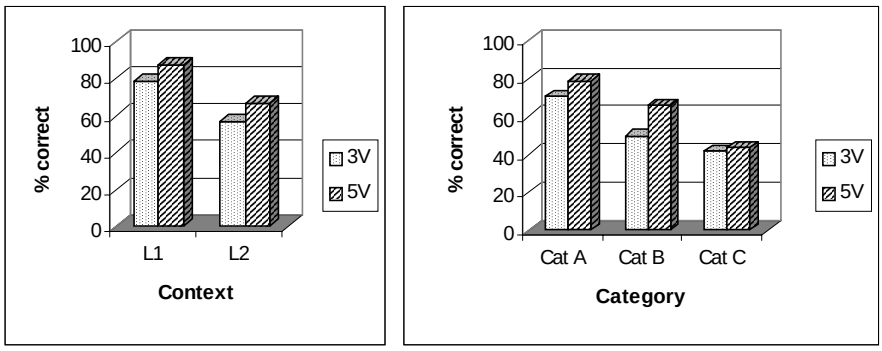

Figure 31. Percentage correct scores on the questions about the meaning of the affixes in pseudo-words for both subgroups with regard to the different sub-tests (on the left-hand side) and the affix categories in the L2 context (on the right-hand side).

4.2.5.2.1 Subject groups

The overall difference between the subgroups was not significant, neither in the test involving the language context ($F[1,32]=3.1$; $p=0.088$), nor in the test involving differences between the affix categories ($F[1,32]=2.1$; $p=0.157$). None of the between-subject variables included in this experiment turned out to be significant at $p<0.05$.

4.2.5.2.2 Categories

The differences between the categories turned out to be significant ($F[2,64]=49.08$; $p<0.001$).

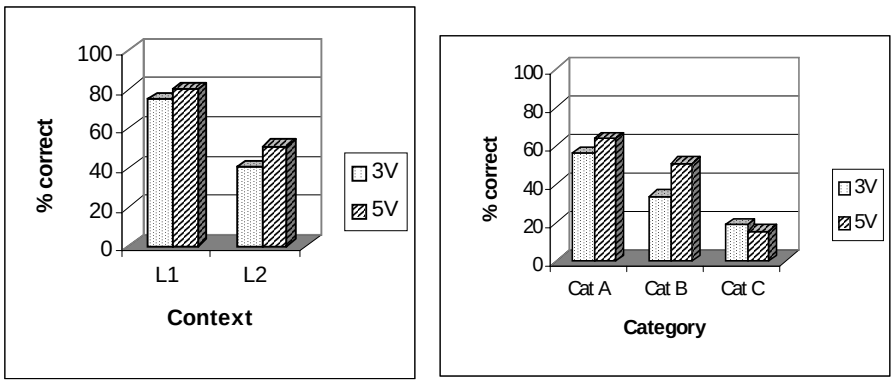

Figure 32. Percentage correct scores on the questions about the syntactic category of the affixes in pseudo-words for both subgroups with regard to the different sub-tests (on the left-hand side) and the affix categories in the L2 context (on the right-hand side).

4.2.5.2.3 Dutch and English sub-test

The scores on the Dutch sub-test were significantly higher than the scores on the English sub-test ($F[1,32]=84.19$; $p<0.001$). Furthermore, a significant correlation was found between a subject's score in the L1 test and the L2 test for both the semantic ($r_{xy}=.64$) and the syntactic ($r_{xy}=.34$) part of the test.

4.2.5.2.4 Syntax and meaning

The additional variable that had explicitly been included in this experiment was the subjects' score on the attribution of the syntactic category of the pseudo-words. Table 6 reflects the co-occurrence of correct, incorrect and blank scores on these variables. This table shows that a larger proportion of correct scores was found for the semantic characteristics than for the syntactic characteristics of the pseudo-words. This difference was significant at $p<0.01$ ($\chi^2 =404.2$; $df = 4$). In both MANOVAs run for the analysis of this experiment (one involving language context and one involving the difference between the affix categories), the difference between scores on syntax and on the meaning of the affix types turned out to be significant ($F[1,32]=14.35$; $p=0.001$ and $F[1,32=28.32$; $p<0.001$ respectively). Furthermore, the correlation between the scores on the syntactic category and the scores on affix meaning was significant at $p<0.05$ in both the English ($r_{xy}=.40$) and the Dutch sub-test ($r_{xy}=.53$).

Table 6. Co-occurrence of scores on the semantic and syntactic section of Experiment 3. The figures in this table represent actual frequencies, not percentages.

		Score syntax		
		Blank	**Correct**	**incorrect**
Score sem	**Blank**	17	8	25
	Correct		477	213
	Incorrect		97	149

4.2.5.2.5 Interactions

A significant interaction was found between the semantic-syntactic variable and the language context ($F[1,32]=7.85$; $p=0.009$. The difference between the syntactic and semantic score was larger in the L2 context than in the L1 context (see Figure 33). None of the other interactions were found significant at $p<0.05$.

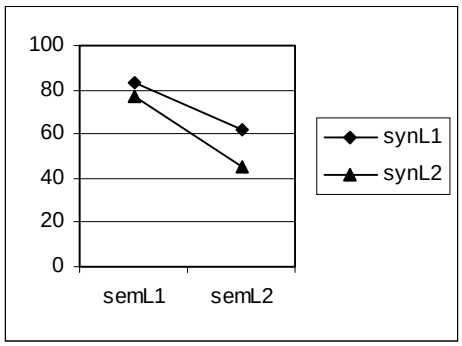

Figure 33. The interaction between the syntactic/ semantic score and language context in Experiment 3.

4.2.5.3 Discussion

Contrary to what had been expected, the difference between the two subgroups was not significant in this experiment. This was not so much due to a small difference in means between the groups (which is obvious from the figures above), as to the high standard deviations associated with these means. For instance, the mean for the semantic score on Category B in group 3V was 50.1 (see Figure 31) with SD=22.2. Apparently, these figures are strongly affected by individual differences. This, together with the significant correlation between the subjects' scores on the L1 test and the L2 test, seems to indicate that the scores in this experiment are affected by individual differences related to the skill to use morphological types. These differ-

ences could not be attributed to age and stage of L2 learning, since these two factors did not significantly affect the scores.

A significant effect was found in this experiment for the categories of the affixes. The largest percentage of correct scores (for the semantic part of the experiment) was found in Category A, followed by B and C respectively. This is perfectly in line with what had been expected. Furthermore, a major proportion of the "semantic" scores in Category C was left unanswered (50% blanks). A difference with the other two experiments is that the percentage of blanks occurred for both sub-groups (53 per cent for the third form learners; 47 per cent for the fifth form learners). Closer inspection of the data revealed that the number of blanks in Category C can largely be attributed to one pseudo-word, *lenksome*, which was massively skipped (97 per cent blanks).

Similar to the previous experiments, the difference between the categories emerged in this experiment, both for the scores on the syntactic categories of the affixes and for the score on the meaning of affix types. Clearly, Dutch learners find it most difficult to interpret morphologically complex L2 pseudo-words based on affix types that do not have a consistent translation equivalent in their L1 (Category C). Moreover, learners make the fewest errors with regard to types that are similar in L1 and L2 in terms of form and in terms of meaning (Category A). The expected interaction of the affix category and the between-subject variables did not occur: the expectation that the largest between-subject difference would occur at Category B scores could not be confirmed.

The differences found between the Dutch and the English sub-test are not very surprising. Obviously, learners are more able to interpret and produce morphologically complex words type-familiarly in their native language than in their second language, due to more fully developed semantic representations of their L1 types. This finding is in line with the expectations expressed in section 4.2.5 above.

The difference between the scores on the syntactic part and the semantic part evidently makes clear that learners from both subgroups have more problems referring to the correct syntactic category of a pseudo-word than referring to the meaning of that word. This, again, is in line with what had been expected. The interaction found between the syntactic and the semantic scores on the one hand and the language context on the other makes clear that this difference is largest in the L2 context. Apparently, learners find it more difficult to reflect on the syntactic category of pseudo-words in the L2 than in the L1. This difficulty is strongest for the affix types in Category C. Again, this is in line with what had been expected.

4.2.6 General discussion

This section summarises the findings for the main variables across the experiments, the subgroups, the categories and the sub-tests, and attempts to account for these findings in terms of the model proposed in the previous chapter.

4.2.6.1 Between-subject variables

An important issue in the model presented in this book is the role of type-form relations as a function of transparency in the development of L2 morphology. In this cross-sectional study, the difference between the subject groups represents the development of morphological acquisition over time. Extra between-subject variables included were the learners' age, the years of formal instruction in English and the amount of exposure to English.

Only in the translation experiment was the overall difference between the subjects found to be significant. This difference, it appeared, was particularly due to a difference in age. This finding was surprising, as the largest between-group difference was expected to occur in more abstract tasks (like in Experiment 3). An explanation for this finding could be that the two subgroups were not very far apart in terms of L2 development. Larger differences may have been found if the distance between subgroups had been larger. A closer inspection of the data, however, does reveal some relevant differences between the subgroups in the other subgroups as well. Firstly, there were significantly more blanks in the scores of the less advanced learners, which indicates that more advanced are more tended to create new words on the basis of morphological types when they are not familiar with a certain type; they have developed the idea of productivity. Secondly, in the comprehension test (the lexical decision task, Experiment 2), there were significantly fewer L1-induced errors in the fifth form than in the third form. Apparently, third-form students more often fail to recognise the morphological type based on the form provided, and more often tend to interpret L2 forms in terms of similar L1 types.

Surprisingly, the effect of the between-subject variables included varied between the experiments. The age factor was only significant in Experiment 1, while the effect of years of formal instruction was only significant in Experiment 2. It had been expected that these factors (and the Exposure Index) would affect the percentage of correct scores in all tests. An explanation for this difference is not obvious; it may either be due to the difference in tasks between the experiments, or to the relatively small samples. Since the individual differences between the subjects appeared to play a major role in these experiments, the latter possibility is most likely.

4.2.6.2 Categories

In all experiments, the largest proportion of correct responses was found in category A, followed by B and C respectively. This clearly indicates that both form-based and meaning-based similarity of L1 and L2 affix types play an important role in the production and comprehension of morphologically complex words in the second language. The more the forms (at the level of lexemes) and the meanings overlap, the more likely learners are to match type and form. This finding is confirmed by the interaction between years of instruction and Category, found in Experiment 2. Development of L2 morphology in terms of the acquisition of morphological types only occurs in Categories A and B, as the types in Category C are not sufficiently consistent between L1 and L2 to allow the acquisition of type-familiarity. Since this effect did not occur in the other experiments, the validity of this observation has to be determined in further research.

Looking at the type of errors made (distinguished in Experiment 1), the data show that very few errors were made in the syntactic category of the affixes in both comprehension and production. This confirms the idea that the types are mostly recognised, but that the real problem lies in matching the English type to the appropriate conceptual representations. Another interesting finding is that most L1-induced errors were found within category C. This makes perfect sense, since interference is most likely to occur in the category where the least consistent and sometimes even confusing relations between form, type and concepts are found.

4.2.6.3 Sub-tests

It is not very surprising that the subjects did generally better in the comprehension than in the production test in Experiment 1. This difference, which can be seen as similar to what Kroll (1993) labelled "translation asymmetry", can simply be accounted for by assuming that learners will have more fully developed semantic forms of their L1 lemmas than of their L2 lemmas. The fact that this effect was not found for the Category A words can be explained by the form-based and semantic similarity of L2 and L1 types in this category that facilitates acquisition and use of these types, leading to more fully developed semantic forms

The overall difference between the number of correct scores in the English and Dutch sub-tests (in Experiments 2 and 3) was evident and was not surprising: subjects can be expected to score better in their native language, as the amount of exposure to the native language is many times larger than the exposure to the second language, leading to more fully developed semantic forms of the morphological types. However, a relatively strong correlation was found between the subjects' scores in the Dutch and the English test. This seems to indicate that it is not only the knowledge of L2 morphology that plays a role, but also the ability to apply type-familiarity in general: subjects that scored well in the Dutch test also scored well in the English test (in Experiment 3, for instance, r_{xy}=0.71). Apparently, the use of language-independent cognitive strategies is an important variable in Experiments testing the performance of L2 morphology.

An interesting effect found in Experiment 2 was that L2 instruction seems to affect L1 performance on the application of morphological types. It may well be that the study of a second language leads to an increased awareness of morphological complexity, regardless of particular affix types and even regardless of the language in which these types occur. In the model outlined in the previous chapter, interlingual co-activation is indeed expected to be bi-directional.

4.2.6.4 Syntax and meaning

In Experiment 1, 50-70 per cent of all scores were correct, implying that the subjects provided the morphologically complex target word intended. 90 per cent of all incorrect scores were morphologically complex words to which an affix had been attached of the same syntactic category as the target word. This implies that for 97 per cent of all answers provided in this experiment the correct syntactic category had been activated. Differences between the subgroups were marginal in this respect. Apparently, the subjects experience no great difficulty in determining and selecting a morphological type with the appropriate syntactic category. The difficulty is obvi-

ously to select the particular affix type that matches the set of conceptual representations specifying the meaning intended.

Experiment 3, which includes scores of the syntactic category, provides a different picture. In this experiment, the scores on the syntactic category are significantly lower than the scores on the semantic category. However, the syntactic scores in Experiment 1 can hardly be compared to those in Experiment 3, as the latter refers to the ability to reflect on the syntactic category of morphologically complex (pseudo) words, which is a different task altogether. Similar to the semantic scores, the category in which the largest proportion of errors occurred was Category C. This, however, cannot be adequately explained in terms of deceptive transparency and similarity to L1 affixes: all of the deceptively equivalent L2 affixes included in this experiment (*-dom*, *ster* -and *-ful*) are of the same syntactic category as their L1 "equivalents". Apparently, a factor that had not been included in this experiment, the subjects' familiarity with the affix types, played a role in this experiment.

4.2.7 Conclusion

The overall picture provided by this exploratory study is that the use of morphological types is clearly affected by the learner's native language. It has been determined that the largest proportion of correct scores can be found for L2 affix types that show most overlap with L1 affix types in meaning, syntactic category and orthographic form. Moreover, even if affix types are dissimilar in terms of form (as in Category B), the overlap of conceptual representations facilitates the use of type-familiarity in production and comprehension tasks. This is further supported by the observation that very few errors were made in the syntactic category of the items in the experiments: morphological types are selected that do have the correct syntactic properties, but problems may occur in determining the semantic specification of the types. L2 forms that can be considered deceptively transparent from an L1 point of view (i.e. in terms of psychotransparency) yielded the largest proportion of incorrect scores. Clearly, psychotransparency is an important condition for the establishment of lexical entries for affix types in the bilingual mental lexicon.

With regard to the development of L2 morphological acquisition, no definite conclusions can be drawn based on this study. There is some evidence that the skill to use morphological types is more strongly dependent on L1 experience and formal instruction than on the exposure to the target language. In addition, the more advanced learners show a greater confidence in the use of morphological types, which makes perfect sense: the more L2 forms have been matched with their types, the more risk the learner is willing to take in guessing/producing new forms. However, these findings did not consistently appear in all experiments. In these experiments, no evidence was found for a developmental distance between comprehension and production. Although generally a larger proportion of correct scores was found in the comprehension task, no effect of any of the between-subjects variables was found on the difference between comprehension and production. It should be noted that this cross-linguistic study only included two subgroups, which were not very wide apart in the developmental process.

In sum, the findings of this exploratory study support the assumption that the use and acquisition of L2 morphology is strongly dependent on the form-based and semantic similarity to L1 morphological types. However, a few remarks are in place to put these experiments into perspective. First, the number of affixes in each category was variable and generally rather limited. In addition, the division of affixes in categories is rather imprecise. Especially Category C represents two distinct sorts of difficulty: L2 affix types that do have a consistent translation equivalent L1, but that are confusing due to cross-linguistic homonymy (like *-ster* / *-ster*), and type-form relations that sometimes lead to correct assumptions about transparency, but are often misleading (low degree of translation equivalence). An example of the latter is the English affix type *–ful*: this affix type had been attributed to Category C, as *–ful* is not always translated by the phonologically similar Dutch form *–vol* and therefore considered "deceptively transparent" (English *painful*, for instance should be translated by Dutch *pijn**lijk*** and not *pijn**vol***). Sometimes, however, the affix types in this category do appear with similar forms in L1 and L2 (English *respectful* can be translated by Dutch *respectvol*). The division into categories does not separately take into account the degree of translation equivalence of the various affix types in Dutch and English. Second, the number of subjects per group was small, and the stages of development represented by the subgroups were not very far apart. Third, the experiments reported here did not take into account some other relevant variables, like productivity and frequency. Consequently, the conclusions drawn here can only be tentative and further research is needed to confirm these findings.

Further studies were devised to investigate the factors that determine the (psycho-) transparency of L2 morphological types. These studies concentrate on the effect of different degrees of translation equivalence consistency in combination with different degrees of productivity. Two methods of investigation were used: a psycholinguistic priming experiment (involving reaction time measurement), reported in 4.3, followed by a written production task of morphologically complex words in L2, reported in 4.4.

4.3 Testing the links between L1 and L2

4.3.1 Introduction

In the previous chapter it has been argued that morphological surface forms are realisations of lexemes, which represent more abstract lemmas consisting of a lemma node that is linked to entities containing semantic (including pragmatic) information, syntactic information and information associating the lemma with a particular language. Based on this assumption, it can be postulated that lexical entries of L2 morphological types share syntactic and semantic information with their equivalent L1 types to the extent that the L1 and the L2 entries overlap in semantic and syntactic characteristics. Furthermore, it has been argued that lexical processes in both production and comprehension are driven by activation spreading. Nodes with a high

degree of activation will spread activation to adjacent nodes. In comprehension, activation spreading takes place from the lexeme to the conceptual representations and in the reverse direction ("activation feedback"). The ultimate consequence of this assumption would be that a high degree of activation of an L1 lemma node would spread activation to the L2 morphological form (through co-activation and activation feedback, see 3.6). This type of co-activation is the subject of the current experiment.

To investigate the occurrence of interlingual co-activation and activation feedback, a priming experiment was conducted[44] in which morphologically complex words in one language were primed by a transparent morphologically complex word in the other language. This situation is graphically represented in Figure 34. The representation in this figure is simplified in that the lemma nodes have been left out; only the overlap of the semantic forms is indicated. It was ensured that the L1 and L2 morphological types that were selected for this experiment overlapped in terms of conceptual representations, but did not overlap in terms of form (i.e. in terms of orthography and/or phonology).

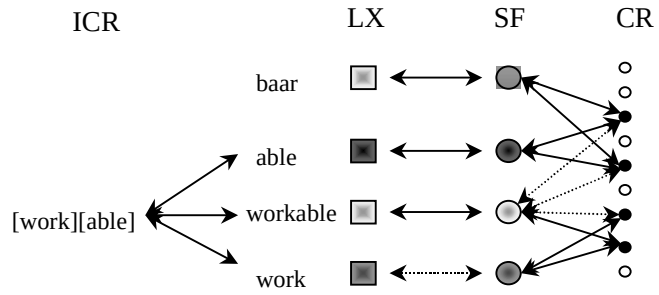

Figure 34 Co-activation of the English affix type -able *after priming of Dutch* -baar.

The experiment was set up in two stages: a pilot study and a follow up experiment. The major objective of the pilot study was to determine which morphological types would be most suitable to be included in this experiment. Secondly the pilot was carried out to determine some technical details, like the maximum response times, some details about the presentation of the target items and priming items and the maximum number of items that the subjects could be confronted with before they started to lose concentration. The pilot will not be reported on separately, but will be referred to in the relevant sections if necessary.

[44] For this experiment, I am greatly indebted to one of my students, Sible Andringa, who carried out the major part of the actual testing and was a great help in devising and analysing the experiment. Of course, any errors in the data presented here remain my own responsibility.

4.3.2 Method

4.3.2.1 Materials

In the pilot experiment, four conditions were used, representing Dutch and English morphological types with (1) similar forms, (2) similar meanings, (3) both similar forms and similar meanings, and (4) neither similar forms nor similar meanings. To avoid priming effects induced by orthographic or phonological similarity it was decided to restrict the affix types in the actual experiment to (3), viz. L1 and L2 morphological types that largely overlap in terms of their conceptual representations, but that bear no form-based similarity (cf. "Category 2" in 4.2.1).

For this category of L1/L2 affix pairs, four different item sets were compiled representing two conditions: a priming condition and a control condition. In the priming condition, a morphologically complex L2 target word was primed by a semantically related transparent L1 prime. For instance, the target word *brightness* would be primed by the L1 type *-heid.* In the control condition, the same target word would be primed by a non-affixed word that was not in any way related to the target item, neither semantically nor orthographically or phonologically. To avoid a practice effect, a split-group design was used. Each group was given half of the items from the control condition and half of the items from the experimental condition, thereby ascertaining that the same target word would be presented to the individual respondent only once. Two affix types were used in each condition: English *-able*, primed by Dutch *-baar*, and English *-ness* primed by Dutch *-heid*. Both affix types are productive in English and in Dutch, although *-ness* is more productive than *-able* (see 4.4.1). The degree of translation equivalence of both pairs turned out to be high, though the translation equivalence of *-baar*/*-able* (93%) is higher than *-heid*/*-ness* (52%; also see 4.4.1). Finally, it was ensured that the focus of attention was on the meaning and category of the affix type by avoiding additional semantic priming between the stems of the primes and the targets. For instance, the English word *thinkable* was primed by Dutch *regelbaar*, thereby focusing on the possible co-activation of the affix types (*-able* and *-baar*) and avoiding cross-linguistic semantic priming of the stems (*think* and *regel*).

The priming conditions and the control conditions for both pairs of affix types were divided over the two subgroups as represented in Table 7.

Table 7. Experimental conditions in the Priming experiment

Subgroup	Prime		Control	
A	*-able* target set 1	*-ness* target set 3	*-able* target set 2	*-ness* target set 4
B	*-able* target set 2	*-ness* target set 4	*-able* target set 1	*-ness* target set 3

4.3.2.2 Items and controls

Each target set consisted of ten items that were controlled for frequency; only low-frequency items were selected to avoid effects of item-familiar word recognition. A t-test was administered on the COBUILD frequency of all item sets to verify this. To control the number of syllables and word length between the experimental and con-

trol conditions, the same stem was used for corresponding pairs in the two conditions. For instance the prime-target pair *regelbaar – thinkable* (*-able*, target set 1) was given the corresponding control pair *regelneef – thinkable*. Some examples of the items and the controls are given in Figure 35. This figure also shows the division of the targets over the groups and the item sets. The full target sets have been included in Appendix 2.

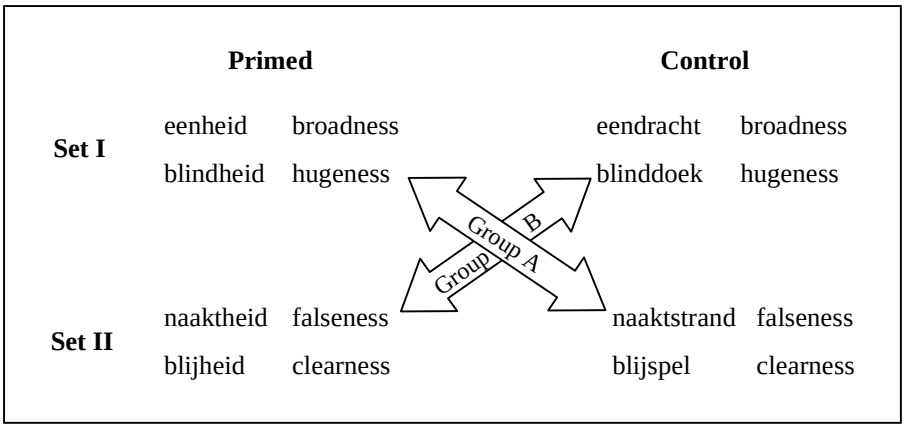

Figure 35. An overview of the design: some examples of items representing the primed condition and the control condition for the -ness *affix types in this experiment.*

4.3.2.3 Fillers

The total number of items in each subgroup was 240, including the practise items. The total number of words was the same as the number of pseudo-words in each group (120). The four conditions were represented by 40 items; 200 fillers were added to each test. 40 fillers were added that were pseudo-words representing the same affix types as included in the conditions. Half of the remaining 160 fillers (80 words and 80 pseudo-words) were morphologically complex, the other half was monomorphemic. All items, however, were polysyllabic.

4.3.2.4 Subjects

The subjects were advanced learners of English in their last phase of their study of English, 43 from the University of Groningen and 17 from the University of Amsterdam. The high level of L2 learning was chosen to ensure that all subjects had acquired the morphological types used in this test. The subjects were randomly assigned to one of the subgroups. None of the students participating was involved in any way in this experiment or any of the previous experiments, and they were not informed about the objective of the experiment. To encourage students to participate in the experiment, four book tokens for 25 guilders each were made available for a lottery in which all subjects participated.

Three subjects had to be excluded due to very slow response rates (>2000 ms) or extremely high error rates (50 per cent in one of the conditions).

4.3.2.5 Procedure

The item files for both groups were scrambled using the "block size" parameter of Dmastr Display System[45]. This resulted in blocks (size 10) containing an equal number of items from all conditions. The program then randomly ordered the items within each block and then finally randomly ordered the blocks themselves. This procedure guarantees an even distribution of conditions across the experiment: in this way it is impossible that most of the items in one condition occur in one half of the experiment and most items of another condition in the other half.

The experiments were conducted in quiet rooms at the University of Amsterdam and the University of Groningen. In both rooms the same computers were used (HP 386sx) with an SVGA colour screen. The subjects were randomly attributed to Group A or Group B. These groups were equally divided over the two computers that were used. The subjects received some instructions on the screen and orally by one of the experimenters. The experimenters were given specific instructions for this purpose (see Appendix 2). After going through a series of twenty trial items the experimenter asked the subject whether everything was clear. If this was not the case, the subject was given the same set of trial items again.

Before the prime was shown on the computer screen, a fixation point (asterisk) was placed at the centre of the screen during approximately 750ms[46] and subsequently replaced by the prime (750ms) and the target (750ms). To accomplish strong activation of the affix, the affix was previewed for 60ms before the start of the timer (SOA prime). The cut-off points were set to 300ms and 2000ms for the fastest and slowest reaction times. Subjects with an error rate larger than 50 per cent in one of the conditions were rejected.

The subject's preferred hand was always associated with the YES response button, the other hand with the NO response button. The experiment was self-paced; the subjects were not able to stop the program or to change the pace. No feedback concerning reaction times or correctness of responses was given during the experiment. After completion of the task (which took approximately twenty minutes), the resulting data file was saved for later analysis.

4.3.2.6 Analysis

A 3-way MANOVA was run, with the test condition (primed and control) and the affix type (*-ness* and *–able*) as within subject factors, and group (A and B) as a non-repeated factor. The effect of the item sets (I and II) could not be included in this

---

[45] The Dmastr Display System ("laboratory software for mental chronometry") v2.0 was used for all stages of the experiment, from compiling the item files to the statistical analysis. This software was made available by the University of Arizona, Tucson, Arizona. I am grateful to the writer of this software, Kenneth Forster, for his useful e-mail assistance to using this program.

[46] The timing in the Dmastr program is based on the refresh rate of the screen. The parameters in the program refer to ticks of the video clock. One clock tick resembles 16.67ms for the type of screen that was used in this experiment.

analysis, because the item sets I and II represent different items for each affix type. To test the possible effect of the item set, a separate ANOVA was run.

4.3.3 Results

The results of the MANOVA with the group as a non-repeated factor are represented in Table 8. A summary of reaction times per word and the matrices of all factors in the experiment have been included in Appendix 2. The overall analysis showed significant effects of the affix pairs and an interaction of group and condition. The main effect, primed vs. controlled condition, was not significant.

Table 8. MANOVA table of the priming experiment with Group as the non-repeated factor. The factors included are the condition (A), the affix (B) and the group (C).

SV	DF	F	P
C	1	3.43	0.0807
S	18		
A	1	0.44	0.517
AC	1	5.43	0.0316
S*A	18		
B	1	43.55	.340E-05
BC	1	0.01	0.916
S*B	18		
AB	1	0.30	0.591
ABC	1	1.39	0.253
S*AB	18		

Condition
The overall difference in scores of the priming condition and the control condition (Figure 36) turned out not to be significant at $p<0.05$.

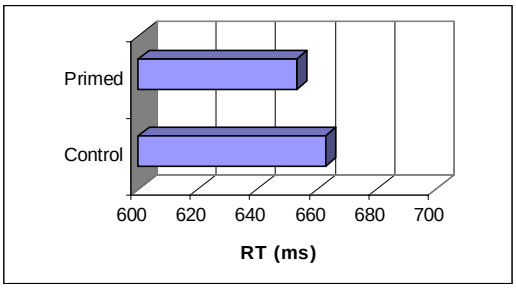

Figure 36. Overall differences between primed and control condition (not significant at p<0.05)

Affix

The overall difference between the reaction time on the *-able* words and the *-ness* words was significant at $p<0.05$. The reaction times on *-able* were faster than the reaction times on *-ness* (see Figure 37). However, this difference did not affect the priming effect, as the interaction between the condition and the affix type was not significant.

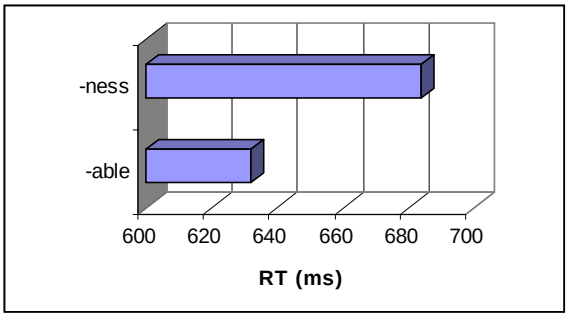

Figure 37. The overall difference in reaction times between the scores on -ness *and* -able.

Group

The overall effect of the group was not significant. However, the interaction between the group and the condition turned out to be significant (see Figure 38). Since each group had been given a different item set, further analyses were done on the differences between the groups and on the differences between the item sets. Of all subjects personal data had been recorded on their main specialisation within English (linguistics, literature or historical language and literature), their age (19-66; mean=24.1), the number of years of studying English at university (3-7; mean=3.9) and their sex. Firstly, none of these variables significantly affected the priming effect. Secondly, the two groups did not differ with regard to any of these variables (Levene test for homogeneity of variance yielded no significant results at $p<0.05$ for the subgroups for any of these variables). Finally, Q-Q plots revealed that both groups showed the normal distribution.

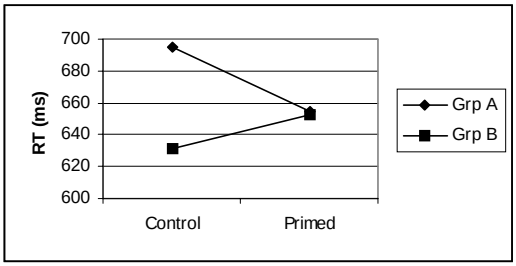

Figure 38. Interaction between the condition and the group (data set).

Item set
The item sets were numbered I to IV: I and II for the different item sets in the *-able* sub-test, and III and IV for the different item sets in the *-ness* sub-test. These sets were entered in an ANOVA with reaction times. The overall difference between the item sets was significant (F[3,79]=6.99; p=0.0003). A Scheffé post hoc analysis, however, revealed that the difference was due to a difference between the affix types: homogeneous subsets were found for Set I, II, III and III and IV. No differences were found between the item sets with regard to the string length of the prime, the string length of the target, the lemma frequency of the prime[47] and the lemma frequency of the target.

Error rates
For the error rates, the same pattern was found as for the reaction times. Again, no significant difference was found between the primed and the control condition (F[1,18]=0.51; p=0.486), as is apparent from Figure 39. The only significant main effect was found for the differences between the affixes *-able* and *-ness* (F[1,18]=17.08; p=0.001); most errors were found for the *-ness* words (see Figure 40). This time, no significant interaction was found with group (F[1,18]=0.166).

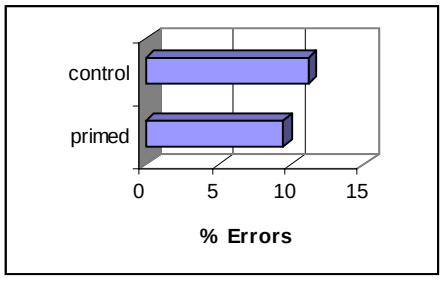

Figure 39. The overall difference in error rates between the primed and the control condition (not significant).

[47] For these data, the 36-mln-word INL corpus was used, available from Leiden University. The subcorpora selected for this purpose were “Varied” and “Newspaper” (25,189,682 tokens together), leaving out “Legal”.

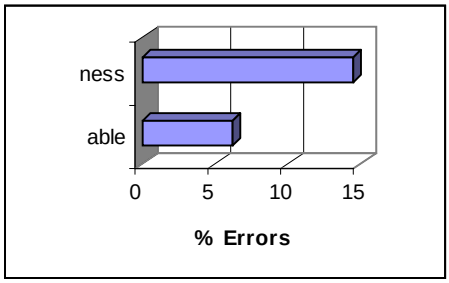

Figure 40. The different overall error rates for the -ness *words and the* -able *words (significant).*

4.3.4 Discussion

The main effect that had been expected, the difference between the primed condition and the control condition, did not occur in terms of reaction times, nor in terms of the percentage of incorrect responses. This implies that the occurrence of interlingual co-activation, postulated in the model, could not be demonstrated by this experiment. The reason why this effect did not show is a matter of speculation: perhaps the groups were not sufficiently large, the level of L2 proficiency of the subjects not high enough or the effect not strong enough to come out in a Lexical Decision Task. An external factor may have interfered (see below).

The difference found between the two affixes, both in terms of reaction times and in terms of errors had not been expected. It was particularly striking that the fastest reaction times and the fewest errors were found for *-able*, while the affix type *-ness* is much more productive than *-able*. Probably, this is related to the difference in an L1 effect (translation equivalence) that had not been included in this experiment (see section 4.4).

The most puzzling effect found, however, was the strong interaction between the item sets and the condition. In one group, the priming effect is rather strong (40ms), but in the group, a reversed effect (22ms) partly neutralises the eventual priming effect found. Clearly, some experimental flaw must have affected the experiment, and it seems of the utmost importance to determine the cause. Since each group had been given a different item set (to ensure each subject had to respond to each target item only once), there are two possible causes for this effect: a difference between the groups and a difference between the item sets. The analyses carried out to investigate possibly interfering effects clearly showed that the difference cannot be attributed to any of the between-subject effect that could be traced, like the subjects' age, sex, years of studying English or their main specialisation. Neither could any difference be found in terms of the length of the items or the frequencies of the primes or the targets between the group scores or between the item sets. Both the groups and the item sets appeared to be perfectly heterogeneous, and in regard to the way the subjects were randomly attributed to the groups, a between-group difference would

indeed be very unlikely. To date, no satisfactory explanation of this awkward result has been found.

4.3.5 Conclusion

In this experiment, the hypothesised effect between the primed condition and the control condition was not confirmed. Since the hypothesised effect was expected to be rather small, an explanation for the non-occurrence of the effect might be found in the fact that interlingual activation feedback is too weak to be demonstrated in a lexical decision task like this. In further research, different methods of investigation should be attempted.

An effect that clearly showed was a difference in reaction time between the two affix types included: the responses to the *-able* words were considerably faster than the responses to the *-ness* words, while fewer errors were made in the responses to the *-able* words. The difference between these two affix types has been further investigated in another experiment (see 4.4).

A significant interaction was found between the two groups and the difference between the primed condition and the control condition. As yet, no satisfactory explanation has been found for this result. It might be worthwhile to conduct a replication study to further investigate the nature and cause of this phenomenon. In Chapter 5, some suggestions to this effect will be made.

4.4 Morphological translation equivalence as a factor affecting the acquisition of L2 morphology

The exploratory study described in 4.2 gave rise to some more detailed questions with regard to the acquisition of L2 morphological types. One of the most pertinent questions concerns the role of the perceived transparency (psychotransparency) of morphologically complex items in the L2 in conjunction with the productivity of L2 morphological types. What are the factors that contribute to the learner's conception of the productivity of morphological types in L2? This is the issue of investigation in this section. Based on a typological comparison of the major morphological types in L1 and L2 (4.4.1), predictions will be made about the degree of translation equivalence of Dutch-English affix pairs, leading to differences with regard to difficulty of learning. These predictions do not assume that more different is necessarily more difficult, but are based on the chance that morphological type in L1 can be translated to a similar morphological type in L2, which is a reflection of the amount of conceptual overlap between these types (see 3.4.3). The second element of the typological comparison concerns L2 productivity. In section 4.4.2 the relative importance of L1-based translation equivalence and L2-based productivity is investigated in a production experiment involving Dutch learners of English from three different levels of L2 proficiency.

4.4.1 A typological comparison of Dutch and English suffixation

In Chapter 3 it has been argued that the consistency of the possibility to translate an L2 affix type by a particular L1 affix type, labelled "morphological translation equivalence", is an important predictor of the psychotransparency of morphologically complex words in L2 (see 3.4.3.3). Morphological translation equivalence is determined by the overlap of semantic and syntactic characteristics of an L1 type and an L2 type. Translation equivalence is not a binary construct, but a continuum: the more syntactic and semantic characteristics are shared by an L1 type and an L2 type, the higher the degree of translation equivalence will be. Since it is not possible to compute the amount of overlap of conceptual representations, translation equivalence is here defined as the chance that a particular morphological type in L1 can be successfully translated into an equivalent morphological type in L2. In this section, a corpus-based approach is described that was used to compute this notion of morphological translation equivalence.

The procedure that was used is very simple: from a representative corpus of English and Dutch random samples were taken of a fixed number of tokens for a range of morphological types. For each morphological type, the number of equivalent translations was counted. For example, consider the seemingly comparable morphological types for nominalisation, the Dutch affix *-heid* and the English affix *-ness*. To determine the extent to which *-ness* is an appropriate translation equivalent of *-heid*, the number of instances were counted where *-ness* could be used to form a morphologically complex English word with the same meaning or function as the Dutch *-heid* word. In the case of this example, it turned out 51 per cent of the Dutch words in *-heid* can be translated by an English word in *-ness*.

4.4.1.1 Corpora

For this typological comparison[48] the lexical databases were used from the Dutch Centre for Lexical Information in Nijmegen (CELEX). The English CELEX database (v. 2.5) is based on the Cobuild corpus, the result of the Cobuild Project of the University of Birmingham. In this corpus, spoken (25%) and written (75%) language has been recorded, from several categories: "broadly general, rather than technical, language; current usage, from 1960, and preferably very recent; naturally occurring" text, not drama; prose, including fiction and excluding poetry; adult language, 16 years or over; "standard English", no regional dialects; predominantly British English, with some American and other varieties" (Renouf, 1987: 2). The English CELEX database contains about 18,000,000 word forms, and provides information about phonology, morphology and frequency. The Dutch CELEX database (v. 3.1) contains 40,000,000 word forms and is based on the corpus from the "Institute for Dutch Lexicography" (INL). Both databases were accessed through an on-line UNIX Telnet connection with CELEX at the Max Planck Institute in Nijme-

[48] For this work I am greatly indebted to one of my students, Esther Bakker, who carried out large parts of this investigation, reported in her MA thesis (see bibliography). Of course, I take full responsibility for the data reported here.

gen. The standard FLEX interface was used to extract data from the database. There are several advantages of using the CELEX lexical database compared to other media. The database reflects current language use and provides frequencies for both spoken and written language, which is more than any dictionary can do. Moreover the database is superior to printed frequency lists currently available (like, e.g. the list by Kucera & Francis, 1967) in accessibility, size and recency. Most importantly, however, the CELEX corpus provides the opportunity of searching for text strings in combination with a wide range of features of orthography, phonology, morphology and syntax by means of complex queries. For English morphology alone more than thirty different features ("columns") are available for each word form, and information is available for word forms and lemmas. It has been calculated that if all columns from the CELEX database were printed for all word forms, a piece of paper would be needed of approximately 5.5m wide and 2.4km long, "so you could probably walk round it in just under an hour" (CELEX manual 1-3). In comparison, the 315,000 entry Random House Unabridged Dictionary of English would need not much more than a strip of 5.5 metres wide and 30 metres long.

For the current study, the following data on morphology were extracted from the database for a set of Dutch and English affixes (see 4.4.1.2). In parentheses are the standard CELEX abbreviations:

Headword (Head)
Headword, reversed (HeadRev)
Spelling number (OrthoNum)
Morphological status (MorphStatus)
Complete segmentation (flat) stems & affixes (Flat)
Complete segmentation (flat) class labels (FlatClass)
Complete segmentation (flat) stem / affix labels (FlatSA)
Immediate segmentation Opacity (ImmOpac)
Noun-verb-affix compound (NVAfffComp)
Derivation method (Der)
Compound method (Comp)
Derivational compound method (DerComp)
Cobuild frequency 17.9m (Cob)
Cobuild 95% confidence deviation 17.9m (CobDev)
Cobuild written frequency 16.6m (CobW)
Cobuild spoken frequency 1.3m (CobS)

Not all of these columns were needed for the selection of all the affixes in the experiment, but these are the columns that were used for both the Dutch and the English database.

The Headword gives the full orthographic form of the resulting lemmas. This column was selected for feedback only: in this column the orthographic form of all lemmas is listed.

The Reversed Headword was selected to enable sorting from right to left. As the focus of this study was to investigate suffixes, all similar words could easily be clustered in this way. For instance, words like, *dogcatcher*, *cowcatcher* and *fly-*

catcher will appear next to each other in a reversed alphabetical list (*rehctacgod*, *rehctacwoc*, *rehctacylf*, respectively). This was necessary for some manipulations that had to be done on the resulting list of lemmas.

The spelling number was included to exclude double occurrences of spelling variants in the lexicons. The output was restricted to spelling nr. 1. Without this restriction, all English lemmas containing the *-ise* morpheme (not only in the *-ise* lexicon itself, but also in cases like *-iser* in the *-er* lexicon) would occur twice. This way, all British and American English spelling variants would be merged and represented only once.

The morphological status provides information about the morphological complexity of the words in the database ("C" for morphologically complex lemmas). For all affixes, this was restricted to "C" only.

The "complete segmentation stems and affixes" renders the surface form plus its flat segmentation (un+like+ly+ness). This column was used for the actual selection of the affixes. The affixes were selected by defining a particular matching string, which would be applied to the "complete segmentation" feature to filter the database output. For the affix *-ness*, for instance, the matching string *%+ness* would ensure that only the lemmas ending in *-ness* were selected that contain the *-ness* affix (*witness*, *baroness*, *harness*). In the string *%+ness*, *%* stands for any number of preceding characters.

"Complete segmentation class labels and stem / affix labels" was selected to check and limit the selection of lemmas. The stem-affix labels provide the number of stems and affixes in the word form (for instance "ASAA" for *unlikeliness*). If this is confined to SA, only those lemmas are selected that consisted of precisely one stem and one affix. The class labels could be used to further confine the lemmas selected. For the selection of *-able* lemmas, for instance, the class labels were restricted to Vx, thereby limiting the resulting lemmas to those that have a verbal stem only.

"Immediate Segmentation Opacity" was included to provide information about the semantic transparency of the lemmas concerned. However, this feature was not used as a limitation, as translation equivalents should be defined as the percentage of successful translations out of all possible morphologically complex lemmas, regardless of their semantic transparency.

The methods of analysis ("Noun-verb-affix compound" "Compound", "Derivational compound" and "Derivation") were selected to avoid occurrence of compounds that would double the presence of particular lemmas. This was particularly necessary for Dutch lemmas, as unlike most English compounds, Dutch compounds are spelled as one string of letters without any spaces. Due to this, a considerable number of morphologically complex forms would occur more than once in the resulting lexicons. For instance, the morphological complex target word *drinker* would re-occur in forms like: *bierdrinker*, *theedrinker*, *koffiedrinker*, *gelegenheidsdrinker*, *kwartaaldrinker*, *sprirusdrinker* and *probleemdrinker*. An English example was found for *catcher*: *flycatcher*, *dogcatcher*, *cowcatcher* and *oystercatcher*. The Dutch lemma *werker* even re-occurred 77 times, some of which recursively: *werker*; *bankwerker*; *machinebankwerker*. The same problem occurred in English for prefixed words; many morphologically complex lemmas containing the prefix *un-*, for

instance, recur elsewhere in the database without that prefix. However, a simple restriction to FlatSA column, limiting the selected lemmas to one stem and one affix would imply a loss of data. For instance, the Dutch corpus contains many phrasal verbs that are spelled as one letter strings. In the corpus the prepositions in these strings are tagged as "S" (for example: *indringer* is tagged as SSA: *in+dring+er*). Omitting all SSA lemmas would lead to the loss of all morphologically complex lemmas that have a phrasal verb as their base. The information about the possibility of different analyses in CELEX, however, creates an opportunity to solve this problem. The first type of compound analysis that is included in CELEX is the noun-verb-affix compound. This category contains all compounds that can be analysed as a nominal stem plus a verbal stem plus an affix. The analysis of many words of this type is ambiguous, and all analyses have been included in CELEX. For some compounds, more than three analyses are added to the database. The selection of this type depends on further analysis of compounds. Pure compounds should not be selected, as these will always lead to combinations of stems that will have been selected anyway, like *mapreader*, which overlaps with *reader*. Therefore, "Comp" was restricted to "N". Derivational compounds, on the other hand should be selected as these can only be formed in combination with a (derivational) affix. For instance, the word *cliffhanger* cannot simply be analysed similar to *coathanger* ([coat][hanger]) as a right-headed compound of the "isa" type ([cliff][hanger]). The analyses of these words in CELEX are [[coat],[[hang],[er]]] and [[cliff],[hang],[er]] respectively. This shows that words like *cliffhanger*, as opposed to *coathanger* should be included, as there is no obvious overlap between *cliffhanger* and *hanger*[49]. A third type of analysis that was included concerned those compounds that do not comply with any of the previous analyses, but that must be seen as derivations. An example of this type is *proofreader*, which has been tagged in CELEX as a compound verb (*to proofread*) plus an affix. Setting "Der" to "Y" (Yes) covered this inclusion of this type".

Finally, frequency data were included for all affixes selected. For the English affixes, the frequencies were added from the 17.9 million Cobuild corpus (spoken plus written frequency) and for the spoken and the written corpora separately. In addition, the column CobDev was added enable checking of the reliability of the frequencies. This is the deviation figure for the lemma frequencies that were estimated based on word form frequencies[50]. Similar columns were added to the lexicons containing the Dutch affixes. This time, the frequencies are based upon the 40,000,000 Dutch INL corpus. The Dutch corpus, however, does not contain frequencies of spoken language. Both the Dutch and the English databases contain many lemmas

[49] It might be argued (see, for instance, Lieber, 1980) that for these words an underlying verb *to cliffhang* must be assumed, which does not occur as a surface form.

[50] For strings with a frequency > 100, disambiguation these strings into lemmas was not done by hand. In those cases, as estimation of the distribution of lemmas for that particular string was made on the basis of the manual analysis of 100 strings. For these entries the deviation figure provides essential information about the reliability of the frequencies. See the CELEX manual for further details.

that have been given frequency 0. These words were listed in the dictionaries that had been used in compiling the database, but were not actually found in the corpora. Since the purpose of this investigation was to investigate the actual use of words in a language, rather than what is in a dictionary, a selection restriction was added to the queries that limited the output to lemmas with frequency greater than 0.

Table 9. Example of database selection from CELEX. In this example the following columns were selected: Headword lowercase alphabetic (Headlow); Morphological status (M); Immediate flat segmentation into Stem and Affix labels (FlatSA); Complete Flat segmentation stems and affixes (Flat); Overall Cobuild frequency (Cob); Written frequency (CobW); Spoken frequency (CobS); Immediate Opacity (I).

SHOW

HeadLow	M	FlatSA	Flat	FlatClass	Cob	CobW	CobS	I
acceptable	C	SA	accept+able	Vx	518	472	46	N
accountable	C	SA	account+able	Vx	82	80	2	N
achievable	C	SA	achieve+able	Vx	9	9	0	N
actionable	C	SAA	act+ion+able	Vx	6	5	1	N
adaptable	C	SA	adapt+able	Vx	49	47	2	N
adjustable	C	SA	adjust+able	Vx	28	27	1	N
admirable	C	SA	admire+able	Vx	190	176	14	N
adorable	C	SA	adore+able	Vx	31	31	0	N
advisable	C	SA	advise+able	Vx	76	73	3	N
agreeable	C	SA	agree+able	Vx	197	195	2	N
								V

START GOTO ZOOM HIDE COUNT PRINT SAVE QUERY

Page: 1 (2) Columns: 9 (9) Tempo: 10 Count: 465 ^

4.4.1.2 Selection of affixes

The main aim of this study was to determine the translation equivalence to be tested in an empirical study. To test the impact of translation equivalence on the acquisition of L2 morphological types, affix pairs had to be selected that represent a range of the variables involved. For reasons of feasibility, the selection was restricted to affix types that are traditionally called “derivational”, even though the model presented in Chapter 3 does not make a principled distinction between derivational and inflectional morphology. The following affix types were included in this investigation: *-able*, *-dom*, *-ee*, *-er*, *-ful*, *-hood*, *-ing*, *-ish*, *-ity*, *-less*, *-like*, *-ment*, *-ness*, *-ship*, *-some* and *-ster* from the English corpus, and *-achtig*, *-baar*, *-dom*, *-ement*, *-er*, *-heid*, *-ing*, *-iteit*, *-lijk*, *-loos*, *-schap*, *-sel*, *-ster*, *-vol* and *-zaam* from the Dutch corpus. This selection is representative in degrees of productivity, form-based similarity and dissimilarity between equivalent Dutch and English types, and includes some affixes that can be expected to lead to formations that are "deceptively transparent" from a cross-linguistic perspective. This was determined by looking at the data of an earlier study, comparing form-based and semantic similarities of Dutch and English

types (Lowie, 1991). The expectations with regard to the productivity of the English types were based on the typological descriptions by Bauer (1983) and Marchand (1969). At a later stage, this initial selection was narrowed for various reasons. These reasons are discussed below, in 4.4.1.4, where also the final selection of affixes is presented.

4.4.1.3 Procedure

Some lemmas in the resulting CELEX database queries (labelled "lexicons" within CELEX) occurred more than once, due to distinctions other than the ones selected. The selected corpora in CELEX were exported to an external file, using the "distinct" feature, thereby ensuring that "identical" forms were exported only once. These export files were transferred from CELEX using the ftp facility and were converted and retrieved into a spreadsheet application (Microsoft Excel 7.0) for further processing.

From the English and the Dutch databases, a random selection was made of all the resulting lemmas for each of the affixes. Of each affix type, thirty lemmas were randomly selected. This was done by having the spreadsheet attribute a random number to each row and then sorting the list by this column. The first thirty items were taken from the resulting random list. Although a set of thirty random items could thus be compiled for most affix types, for some affix types less than thirty lemmas were found to meet all selection criteria. This was the case for Dutch *-ement*, both Dutch and English *-dom* and English *-hood*, *-like*, and *-some*. In these cases, no further selection was made, but all the resulting lemmas were used, although there were less than thirty cases. Next, all the lemmas in these thirty-line corpora were translated to the other language. For all affixes in both languages, a morphologically complex translation was chosen if this was possible, even when this was not the most frequent translation. The only restriction to this was that the translation had to be one of the alternatives given in the van Dale dictionaries Dutch-English and English-Dutch (Martin & Tops, 1984), i.e., for those lemmas that were included in that dictionary. For lemmas that were not in these dictionaries, the target word was checked with either the Random House Unabridged Dictionary of English or the Van Dale Woordenboek der Nederlandse Taal. Furthermore, if possible, a word consisting of the same stem as that of the morphologically complex word in the source language was chosen, and, if possible, a preference was given to translations containing "an affix similar in form to the affix in the source language" (Bakker, 1996: 39). If a word in the source language could be translated by a morphologically complex word in the target language, this would be considered as a morphological translation equivalent. For each of the different affix types that yielded a translation equivalent for a particular form in the source language the number was listed and percentages were calculated. From the thirty-word corpus representing the Dutch affix type *-achtig*, for instance, ten times an English translation was possible of the form *-like*. It was thus stated that the translation equivalence of the cross-linguistic affix pair *-achtig*/*like* was 33 per cent. The translation equivalence (TEq) of the reverse pair, with English as the source language and Dutch as the target language (*-like*/*-achtig*) yielded a score of 24 per cent. The translation equivalents that are the result of this selection procedure must be regarded as the maximum transla-

tion equivalence of a particular affix pair. As the procedure and the criteria were identical for all affix pairs in the test, the lenient selection procedure could not in any way affect the outcome.

The resulting lists of affix pairs were ordered by the source language affix. For each source affix a list was compiled of possible translation equivalents, ordered from strong to weak degrees of translation equivalence. Phonological and orthographic varieties of the same affix (like, for instance, *-able*, *-ble* and *-ible)* were considered as representing the same morphological type and were taken together.

The productivity of each affix was calculated by dividing the number of hapaxes by the total number of tokens representing the same affix. The affix *-ness*, for instance, is represented by 1353 types, which have an average frequency of 14.9. The total number of lemmas in the corpus ending in *-ness* and complying with the other selection criteria is 20,179. The number of hapaxes is 209. Therefore, it can be stated that the productivity of this suffix based on hapaxes is n_1/N= 0,010357. See 2.5.1 for a discussion of productivity. [51]

4.4.1.4 Some problems and choices

The first problem that occurred during the selection of the databases from CELEX was that the Dutch and the English corpus turned out not to be fully compatible. The data in the Dutch corpus, for instance, had not been tagged for opacity and did not contain the spoken frequency information. Therefore, the information on spoken corpus frequency and opacity were not included in the analyses. A more serious problem in this respect was that the encoding procedures that were used for the Dutch and the English databases turned out not to be identical. Whereas for English lemmas of the type SA, like *twister*, the Compound analysis was given the value "N", similar lemmas in the Dutch database had been given "Y" for this same feature (*zaaier*). Hence the selection restriction "Comp=N" turned out not to work for the Dutch lemmas. Moreover, the Dutch corpus contained many doubtful or awkward tags. The headword *wijnproever* (wine taster), for instance, was labelled "NNx" in the FlatClass column, which would indicate that this word does not contain a verbal stem. This problem was solved by adding some specific restrictions for some of the affixes in the Dutch database[52] and by manually checking the output file on overlapping occurrences. For this purpose, the file was sorted by the HeadRev feature.

---

[51] The corpus use of "type" should not be confused with what has been defined as a "morphological type" in this study: a corpus type is one particular word form that may occur a number of times in the corpus. Each individual occurrence of a corpus type is conventionally labelled a "token".

[52] For example, the selection criteria for Dutch *-er* were as follows: OrthoNum=1 AND Flat=%+er AND Cob>0 AND (FlatClass=Vx OR FlatClass=AVx OR FlatClass=BVx OR FlatClass=PVx OR FlatClass= VVx OR FlatClass=xAx OR FlatClass=xVx). Some examples of the results are Vx: denker; Avx: dwarsligger; BVx: aangever; PVx: opschepper; VVx: zweefvlieger; xAx: versneller; xVx: behanger.

A problem which occurred as a result of the selection procedure was that homonymous morphological types could not be distinguished. An example of this is the different types underlying the words *mouthful* and *doubtful*. The first type expresses a quantity, whereas the latter expresses quality or a state. It is obvious that, although these two types may have many overlapping lexical and conceptual characteristics, they should be considered as different types. In this particular case, the same occurs with Dutch *-vol*. Other, more subtle differences can be found in the *-er* affix in both languages, which represents both agents (*reader*) and instruments (*tranquilliser*). The distinction of these types can be hard to make and for the current study making these subtle distinctions was not considered feasible. One reason is that no extensive description of all morphological types is available for Dutch and English.

4.4.1.5 Results

In the table below, the results of this investigation are represented for translation equivalents that reached a value greater than or equal to 20 per cent.

Table 10. English translation equivalents of Dutch affix types, % ≥ 20.

Dutch suffix	English Translation Equivalent	% of types	Example
-achtig	-like	33	lenteachtig-springlike
-achtig	-ous	20	monsterachtig-monstrous
-achtig	-y	20	regenachtig-rainy
-baar	-able/-ible	93	bereikbaar-approachable
-dom	-ity	24	adeldom-nobility
-ement	-ment	44	amusement-amusement
-ement	-ion	22	isolement-isolation
-er	-er	77	krasser-scraper
-heid	-ness	51	klamheid-dampness
-ing	-ion	47	beperking-restriction
-iteit	-ity	93	formaliteit-formality
-lijk	-able	27	draaglijk-bearable
-lijk	-al	27	natuurlijk-natural
-loos	-less	83	stemloos-voiceless
-schap	-ship	43	leiderschap-leadership
-vol	-ful	30	betekenisvol-meaningful
-zaam	-able	23	vreedzaam-peaceable
-zaam	-ive	23	werkzaam-effective

Table 11. Dutch translation equivalents of English affix types, % ≥ 20.

English suffix	Dutch Translation Equivalent	% of types	Example
-able	-baar	57	readable-leesbaar
-able	-lijk	23	presumable=vermoedelijk
-dom	-heid	29	wisdom-wijsheid
-dom	-dom	21	dukedom-hertogdom
-ee	ge-de	31	addressee-geadresseerde
-ee	-er	28	lessee-huurder
-er	-er	70	abstainer-onthouder
-ful	-ig	30	powerful-machtig
-ful	-lijk	20	baleful-verderfelijk
-ful	-vol	20	hopeful-hoopvol
-hood	-heid	35	adulthood-volwassenheid
-hood	-schap	35	brotherhood-broederschap
-ing	past part. -d	20	enterprising-ondernemend
-ish	-s	43	slavish-slaafs
-ish	-achtig	20	blueish-blauwachtig
-ity	-heid	70	security-zekerheid
-ity	-iteit	23	passivity-passiviteit
-less	-loos	60	bottomless-bodemloos
-less	on-	23	baseless-ongegrond
-like	-lijk	41	businesslike-zakelijk
-like	-achtig	24	hornlike-hoornachtig
-ment	-ing	83	assessment-beoordeling
-ness	-heid	90	briskness-vlugheid
-ship	-schap	60	friendship-vriendschap
-some	past part -d	27	lightsome-lichtgevend
-some	-lijk	21	troublesome-zorgelijk
-ster	-er/-eur	67	trickster-oplichter

4.4.1.6 Discussion

The approach taken here is not without problems. For instance, it was sometimes difficult to decide on the translation of the source words. Especially in the case of productive affix types, it was sometimes difficult to choose between equivalent alternatives without any context. An example of this is the Dutch word *wasachtig*, which can be translated by either *waxlike* or *waxy*.

It clearly appeared in these results that the translation equivalents are not always equally productive in both directions. This asymmetry is most obvious for the relations between *-iteit* and *-ity* and between *-heid* and *-ness*; *-iteit* can almost always be translated by *-ity*, but *-ity*, like *-ness*, is almost always translated by *-heid* (see Figure

41) The asymmetry between *-able* and *-baar* is most striking. *-Able* is a translation equivalent of a wide range of Dutch affixes: *-lijk*, *-zaam*, *-baar* and *-vol* (see Figure 42). This seems to point to a wide range of meanings covered by the affix *-able*.

It has been pointed out in 4.4.1.4 that homonymous affix types could not be distinguished other than by manual selection. It should thus be taken into account that some of the figures presented here may be affected by homonymous types. A global investigation on two of these types, instrumental *-er* and quantitative *-ful* revealed that both of these types had higher translation equivalence values than the types they were embedded in. Whereas the overall translation equivalence of the pair *-ful*/*-vol* is 20 per cent, the translation equivalence for the type of *-ful* referring to a particular quantity (*plateful*, *pailful*) amounts to a translation equivalence of about 60 per cent.

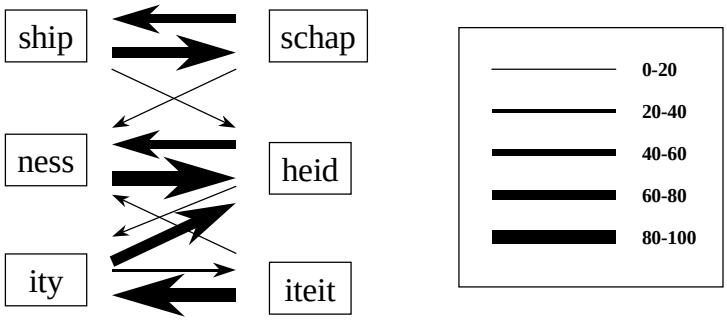

Figure 41. Dutch and English morphological translation equivalents for nominalisation. The width of the arrows reflects the degree of translation equivalence (see inset).

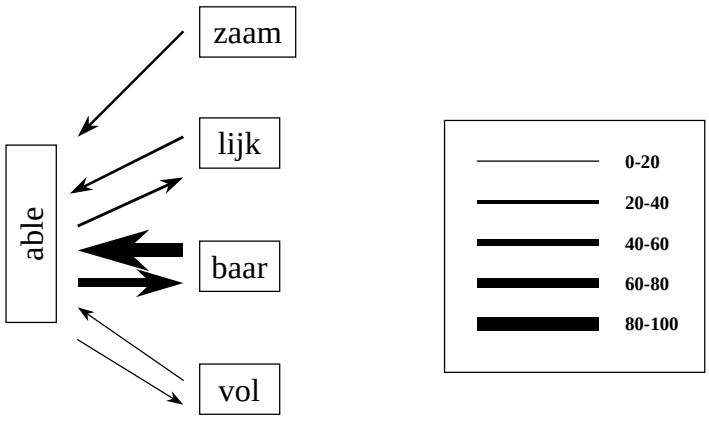

Figure 42. Dutch and English morphological translation equivalents English morphological type -baar. *The width of the arrows reflects the degree of translation equivalence (see inset).*

Translation equivalence thus determined expresses for each affix pair listed above the chance that the affix type in the target language could translate the affix type in the source language. Similar to productivity, learners can be expected to become sensitive to these chances due to exposure, leading to consistent co-activation. If this is indeed the case, translation equivalents computed in this way might be an accurate predictor of the learner's acquisition of morphological types, as was argued in Chapter 3. However, it will be obvious that translation equivalence cannot straightforwardly predict the areas of difficulty learners experience in the acquisition and use of morphological types. Like all aspects of cross-linguistic influence, translation equivalence must be seen as a factor that interacts with numerous other factors to determine the influence of the learner's first language on the acquisition of a second language. Cross-linguistic influence interacts with the learner's age, stage of learning, sociolinguistic factors et cetera. Not all these factors could be included as independent variables in the study presented here, but will be controlled as far as possible. There are three other factors, however, that can be expected to interact with translation equivalence (see Chapters 2 and 3): transparency, frequency and productivity. Translation equivalence itself will greatly contribute to the transparency of L2 morphological types, i.e. the psychotransparency of morphologically complex words for learners of a second language. A type that bears much similarity to a type the learner is already familiar with in her L1 will facilitate the acquisition process of that type due to an increase of transparency. On the other hand, morphologically complex words that seem transparent judged by L1 rules, but in fact are not ("deceptive transparency") are likely to impede acquisition of the types concerned. Transparency, we have seen, is a necessary condition for the acquisition of productivity. Hence, it can be expected that transparency interact with productivity in the acquisition of L2 morphology. To explore the relations between frequency, productivity and translation equivalence, an experiment was set up.

4.4.2 Testing translation equivalence

4.4.2.1 Introduction

The corpus study outlined above quantifies the translation equivalence of Dutch and English affix pairs. In the bilingual mental lexicon, morphological translation equivalence represents the amount of syntactic and semantic information that is shared by the L1 and the L2 affix types. Through interlingual activation feedback, it has been argued in Chapter 3, co-activation occurs between the "equivalent" affix types. Consistent interlingual co-activation as the result of translation equivalence can thus be said to represent the L1 influence in the production of (transparent) morphologically complex L2 words. L1-L2 affix pairs with a low degree of translation equivalence will hamper the production of morphological complex words representing that type. The more consistently an L1 affix overlaps with an equivalent L2 affix (i.e. the higher the degree of translation equivalence is), the more facilitation

will occur for the production of morphologically complex L2 words representing that affix type.

The productivity of affix types, also determined on the basis of a corpus study, can be seen as the L2 factor affecting the production of morphologically complex words in the L2. Contrary to the translation equivalence of L1-L2 affix pairs, the productivity of an L2 affix type is solely determined by L2 factors, without any influence of the L1 (see 2.5.1). To determine the degree of productivity of an L2 affix type, a large amount of exposure to the L2 is required. Therefore, the "knowledge" of the productivity of L2 affix type will increase over time.

An experiment was conducted to investigate the role of the morphological translation equivalence and the productivity of L2 types in the production of transparent, morphologically complex L2 words.[53] In this experiment, Dutch learners of English were required to produce morphologically complex words that varied in their degree of translation equivalence and productivity. The experiment consisted of a cross-linguistic translation task and a monolingual gap-filling task in the L2. The sequence of L2 development in the production of morphologically complex words was incorporated in this study cross-sectionally by including three naturally occurring groups representing different levels of L2 proficiency. It was attempted to control the frequency effect in this experiment by exclusively including low-frequency target items.

4.4.2.2 Method

In this experiment, 116 Dutch learners of English from three different levels of proficiency were asked to produce morphologically complex English words in two contexts: cross-linguistically in a translation task and in an L2 gap filling task in which no reference to L1 affix types was made. The target items in this test were all transparent morphologically complex words representing varying degrees of productivity and translation equivalence.

4.4.2.2.1 Subjects

As this study focuses on L2 learners of English, groups of learners had to be found representing different levels of English proficiency. In the Dutch situation, these groups can best be found at secondary schools, as English is a compulsory subject for all pupils there. There are, however, some disadvantages to using these groups. First, there may be internal differences within these groups in terms of age, exposure to English and years of formal instruction in English (see section 4.2). Another limitation is that the lowest levels of L2 proficiency could not be included in a morphological test, as these learners appeared not to deal with the rather abstract tasks in this experiment. In a pilot test, it was found that the lowest level at which pupils could be reliably tested on their written morphological performance was the third form of Dutch VWO (pre-university education). The three levels were third and fifth

[53] I would like to express my gratitude to two of my students, Wil Hamminga and Gudy Buitink, for their assistance in this experiment.

form pupils from secondary schools and first-year students of English[54]. All tests were taken towards the end of the final teaching term.

4.4.2.2.2 Materials and procedure

Dutch-English Affix pairs were selected in such a way that different levels of both translation equivalence and productivity would be included. For translation equivalence the selection of pairs was based on the figures presented in 4.4.1. The productivity of these affixes was computed by dividing the number of hapaxes for a particular English affix by the number of tokens occurring in the Cobuild corpus, using the CELEX lexical database (as discussed in 2.5.1). The results of these calculations are represented in Table 12.

Table 12. Productivity data of a representative set of English affixes. n_1 stands for the number of hapaxes for the affix type found in the corpus; P stands for productivity (as defined in 2.5.1).

Affix	Types	tokens	n1	P	$P.^{10-3}$	Fgem	types/ tokens
DOM	21	3253	0	0	0	154,9	0,006
LING	24	1125	0	0	0	46,9	0,021
FUL(2)	115	15813	4	0,00025	0,253	137,5	0,007
MENT	288	55784	15	0,00027	0,269	193,7	0,005
FUL(1)	147	16426	6	0,00037	0,365	111,7	0,009
EE	40	3972	2	0,00050	0,504	99,3	0,010
ITY	525	45488	44	0,00097	0,967	86,6	0,012
SHIP	77	7125	7	0,00098	0,982	92,5	0,011
ISH	138	10929	12	0,00110	1,098	79,2	0,013
HOOD	27	2548	3	0,00118	1,177	94,4	0,011
ABLE	437	21870	29	0,00133	1,326	50,0	0,020
SOME	31	1205	2	0,00166	1,660	38,9	0,026
ER	1742	83928	157	0,00187	1,871	48,2	0,021
LESS	213	7096	20	0,00282	2,818	33,3	0,030
STER	19	708	2	0,00282	2,825	37,3	0,027
ING	62	1132	6	0,00530	5,300	18,3	0,055
NESS	1353	20179	209	0,01036	10,357	14,9	0,067
LIKE	34	268	5	0,01866	18,657	7,9	0,127

[54] As has been shown in the study described earlier, the learners in these groups do not only differ in terms of L2 proficiency, but may also differ in terms of cognitive maturation. The selection of learners from the highest level of secondary education was an attempt to reduce this effect.

In compiling the Dutch-English affix pairs for this experiment, a compromise had to be worked out between methodological soundness and feasibility. The most reliable results would have been gained if the pairs had consisted of extreme values for translation equivalence and productivity. For instance, a pair like *-vol/-able* (translation equivalence of the pair = 3%; productivity of the target affix = 1.3) could be selected to represent low translation equivalence combined with low productivity. However, hardly any target words could be found to test this pair. Eventually, the seven pairs were selected that are listed in Table 13 and Table 14. The starting point in selecting these pairs was to include two levels of productivity of the target affix and two levels of translation equivalence, both of which would be represented by two different pairs. For productivity, 1-9 was regarded as low and 10 to 19 as high. For translation equivalence, 3-47 was regarded as low and 48 to 93 as high. Unfortunately, only one pair could be found to combine a high level of productivity with a relatively high level of translation equivalence. Furthermore, attributing a translation equivalence value of 43 per cent to the "low" category and 52 per cent to the "high" category is not the most ideal division. However, since the alternative, dividing translation equivalence into three levels, would leave us without a representation of the high/high category, this was the only feasible solution for which sufficient target items could be found.

Table 13. Dutch-English affix pairs selected for the experiment. These pairs represented high and low levels of both translation equivalence (Teq) and Productivity ($P \cdot 10^{-3}$)

affix pair	TEq	$P \cdot 10^{-3}$	Levels
heid/ness	52	10	High/high
baar/able	93	1	High/low
iteit/ity	93	1	High/low
achtig/like	33	19	Low/high
iteit/ness	3	10	Low/high
heid/ity	11	1	Low/low
schap/ship	43	1	Low/low

Table 14. Affixes selected for the different levels of productivity and translation equivalence

		Productivity	
		High	*Low*
Translation Equivalence	*High*	-heid / -ness	-baar / -able -iteit / -ity
	Low	-iteit / -ness -achtig / -like	-heid / -ity -schap / -ship

Two separate tests were devised. Excerpts of the test forms, including the instructions given, have been included in Appendix 3. In the first test, subjects had to

translate morphologically complex Dutch words into English words ("L1 context"). The translation of the roots of the words was given:

Kaalheid (kaal = bald) _____

In the second test, subjects had to fill in morphologically complex words in English sentences ("L2 context"). Sufficient context was provided for the meaning and the syntactic category of the target word to be unambiguously clear. In this test, the roots of the words were given:

Susan is always very radical. Her radical_____ can be rather annoying.

For each affix pair included, six target words were selected (see Appendix 3). A pilot test showed that extremely high proportions of correct scores were found for transparent morphologically complex words with high item frequencies. For some high-frequency items, like *leadership* (F=718) and *friendship* (F=514)[55], the subjects even scored 100 per cent correct. Clearly, these frequencies are high enough for these words to have their own lexical representation, in spite of their transparency. Since the focus of this study was the subjects' ability to use morphological types, rather than their knowledge of vocabulary, only low-frequency target items were selected. In the majority of all cases, the COBUILD frequency was kept below 10. However, for some target affixes (for instance *-ity*) somewhat higher frequencies could not be avoided. A further restriction on the target items used in the test was that they should preferably not contain more than one affix. This was done to avoid morphological environments in which particular affixes are more productive than in neutral contexts. For instance, as Baayen & Lieber (1991) have demonstrated, *-ity* is much more productive than *-ness* if preceded by *-able*. Finally, it was attempted not to include roots ending in a vowel if the affix type to be attached to that root started with a vowel, if this lead to deletion of the vowel in the root (as, for instance in *bribable*). The six target items were equally distributed among two subgroups (data sets *a* and *b*): half of the pupils were given set *a* in the translation task and set *b* in the gap filling task; the other half were given set *b* in the translation task and set *a* in the gap filling task.

Table 15. Data sets in the experiment

	L1 context	L2 context
sub-group 1	data set a	data set b
sub-group 2	data set b	data set a

A pilot test indicated that the order in which items were presented might strongly influence the results. Once a particular affix had been used, the subjects tended to

[55] All frequencies reported here are cumulative COBUILD frequencies for written and spoken language, as included in the CELEX lexical database.

keep on using that particular affix as a "default" affix that was filled in for virtually all items. Although it could not be avoided that subjects were given the impression that the issue of the experiment was affixation, it was attempted to conceal the focus of attention of the experiment, i.e. the seven affix types mentioned above. To attain this, twice as many decoys were added as there were items: each subject was presented with three target items per affix pair in each context (L1 and L2), whereas the number of items that had to be filled in amounted to 63 items per context. To ensure a random presentation of items, the following procedure was followed: first groups were compiled consisting of one target item from each affix pair, complemented with fourteen decoys. Next, both the groups and the items within each group were randomly ordered. Finally, the item file thus composed was presented in two orders, so that eventually eight different forms were used in the experiment. Each subject was given one set of items in each context, containing items from different data sets.

Table 16. Conditions and data sets

	L1 context		L2 context	
	order 1	*order 2*	*order 3*	*Order 4*
data set a	1a	1c	2b	2d
data set b	1b	1d	2a	2c

The subjects were asked to fill in their name and form on the test forms. At the secondary schools the tests were administered by the pupils' own English teacher, who had been instructed to give as little information as possible. The first-year students of English were requested to complete the forms after a one-hour history exam. On the test forms, the subjects were asked to fill out the form seriously and carefully. In the instructions about the correct way of filling out the forms, it was stressed that all gaps had to be filled. To avoid an initial focus on translation equivalents, the monolingual L2 task was always administered before the cross-linguistic task. For both parts of the test (the L1 form and the L2 form), an example was provided, using an affix type that was not in any of the target items. Of the 120 forms filled in, 4 were rejected, either because large parts of the forms had not been filled in or because the task had obviously not been taken seriously.

4.4.2.2.3 Variables

An overview of the major variables included in this experiment is given in Table 17. Not represented in this table are the variables that were included to increase the validity and reliability of the design: data set; order of presentation of the targets; the affix pairs representing the different levels of translation equivalence and productivity and the targets representing the affix pairs.

Although the frequency of the target items had been kept as low as possible, frequency was included in the design to investigate the possible effect of the small differences in frequency that occurred within the affix pairs (see 2.5.6). To this end, the targets representing the affix types were divided into high and low frequency within

each affix pair, regardless of their absolute frequency. In this way, frequency could be entered in the MANOVA.

Table 17. An overview of the variables in the experiment.

Variable	Function in design	Levels
Group	Independent between-subject factor	3
Translation Equivalence	Independent within-subject factor	2
Productivity	Independent within-subject factor	2
Frequency (relative)	Independent within-subject factor	2
Context	Independent (moderator) within-subject factor	2
Use of target affix	Dependent variable	Binary scores, converted into percentage correct
Use of appropriate syntactic category	Dependent variable	Binary scores, converted into percentage correct

4.4.2.2.4 Predictions

In regard to the model outlined in the previous chapters, the following predictions can be made about the effect of the variables included in this experiment and the way they interact.

Group and Context: these variables had predominantly been included to test their interaction with the main factors, translation equivalence and productivity. Obviously, larger proportions of correct scores can be expected at higher levels of proficiency. The overall scores in the L1 context can be expected to be higher due to the nature of the task. In the L2 gap-filling task, the English sentence has to be interpreted, while in the translation task no interpretation is involved. Since the ability to interpret the English sentences can be expected to increase with increasing proficiency, an interaction is predicted between group and context, where less difference between the tasks is expected for the higher levels of proficiency. Furthermore, providing the L1 affix (in the L1 context) may either facilitate or hamper the production of morphologically complex L2 words. It is predicted that this facilitates production where translation equivalence is high, but impedes production where translation equivalence is low. In the latter case, the L1 affix is likely to interfere (interaction context x translation equivalence, see below).

Translation equivalence: It has been argued in Chapter 3 that translation equivalence can be expected to facilitate the acquisition and use of L2 morphological types, as it will contribute to the psychotransparency of morphologically complex L2 words (3.4.3.3). It follows from this that the more consistent the relation is between an L1 affix type and an "equivalent" L2 affix type (i.e. the higher the degree of translation equivalence is), the more facilitation will occur for the production of morphologically complex L2 words representing that affix type. Conversely, low translation equivalence is likely to hamper the use of a particular affix type, as learners will be forced to use an alternative affix type with a higher degree of translation

equivalence. In terms of levels of proficiency, it can be expected that no effect of translation equivalence is found at very low levels of proficiency, since beginning learners have not yet been sufficiently exposed to the L2 to have noticed the co-activation of highly equivalent affix types due to the overlap of conceptual representations. Therefore, it can be expected that the influence of translation equivalence increase with higher levels of proficiency (see 3.3.2.2). In regard to the different tasks included in this experiment, expressed in the variable context, an interaction with translation equivalence can be expected. In the L1 context, the L1 affix is explicitly shown, which emphasises the relation of the L1 affix provided and the L2 affix asked. In the L2 context, the "equivalent" affix is not explicitly shown, and translation equivalence can only implicitly affect production.

Productivity: The resting activation level of more productive affix types can be expected to be higher, as these types occur with many different roots and more often lead to successful type-familiar processing for the comprehension and production of morphologically complex words. A higher degree of productivity will lead to a larger proportion of correct scores in the current experiment. Since the productivity of an affix type is determined by frequency of successful type-familiar processing of morphologically complex words, the facilitating effect of productivity will particularly show at higher levels of proficiency. In regard to the different tasks in this experiment, it can be expected that the strongest impact of productivity is found in the L2 context, as less L1 interference (through translation equivalence) can be expected in that context. The facilitating effects of translation equivalence and productivity can be expected to be cumulative: the highest proportion of accurate L2 production may be expected for affix pairs with a high translation equivalence and a highly productive target affix. However, in contexts where either productivity or translation equivalence is low, one effect will interfere with the other.

Frequency: The frequency of occurrence of morphologically complex words will affect the performance of L2 learners, as high-frequency morphologically complex words will be approached item-familiarly, rather than type-familiarly: the activation level of these items will be higher than the activation level of their constituents. In the case of item familiar processing, it is not likely that an affix is chosen different from the target affix.

Syntactic category: The syntactic category of the affix type is a more global feature than the semantic similarity of affix types, because many specific affix types share the category. Therefore, the overall proportion of correct scores for the syntactic category can be expected to be higher than the scores for the choice of the target affix. It is particularly interesting to consider the cases where another affix was selected than the target affix. It can be expected that for affix pairs with a low translation equivalence a different affix is selected, which forms an L1-L2 pair with a higher degree of translation equivalence.

In sum, the predictions with regard to the main variables in this experiment and their interaction are as follows:

1. A higher degree of translation equivalence will lead to a larger overall proportion of correct scores.
    a. The impact of translation equivalence will be strongest at high levels of L2 proficiency.

   b. The impact of translation equivalence will be strongest in the cross-linguistic task.
2. A higher degree of productivity will lead to larger overall proportion of correct scores.
   a. The impact of productivity will be strongest at higher levels of L2 proficiency.
   b. The impact of productivity will be stronger in the monolingual tasks than in the cross-linguistic task.
3. A higher degree of item frequency will lead to a larger overall proportion of correct scores.
4. The overall scores for the syntactic category will be higher than the scores for the specific affix types.
   a. For affix types with low levels of translation equivalence, more often an alternative affix of the appropriate syntactic category will be chosen, viz. the affix type that forms a higher translation equivalent for the L1 affix.

4.4.2.2.5 Analyses

The answers provided by 116 subjects were scored on two dependent variables, once for the use of the target affix, and once for the syntactic category. Binary scoring (correct=1; incorrect=0) was used for both the affix and the syntactic category: the affix was only scored "correct" if the affix provided was identical to the target item. Spelling variations (*-able* and *-ible)* and spelling errors were ignored. Blank scores were given a separate code (9). Two sorts of analyses were carried out. A MANOVA was applied with context, translation equivalence and productivity as within-subject factors (with two levels all) and group as a between-subject variable (three levels). Furthermore, χ^2 analyses were carried out to investigate the frequency of occurrence of correct, incorrect and blank scores in the test. χ^2 analyses were also used to investigate the effect of the order in which the items were presented and to check the similarity of the data sets. Finally, some correlation analyses were carried out to investigate the effect of absolute item frequency on the proportion of correct scores.

4.4.2.3 Results

First, it was checked whether the data sets (*a* and *b*) and the order of presentation of the target items had affected the results. For this purpose, some cross-tabulations were compiled with these variables and the number of correct, incorrect and blank scores for the target affix. The difference between the data sets appeared not to be significant (χ^2=3.9; *df*=2; p=.137). However, the differences between the two orders turned out to be significant (χ^2=6.7; *df*=2; p=.03). The results for each of the independent variables will be discussed separately below. All data refer to percentage of correct use of the target affixes; analyses involving the syntactic categories of the "incorrect" scores will be reported on afterwards. A first impression of the resulting scores per affix is provided in Figure 43 and Table 18. The differences between the affix pairs representing the levels of translation equivalence and productivity (see

Figure 43) was separately analysed. This analysis showed that the differences between the affixes within the levels were not significant at $p<0.05$ between *-heid/-ity* and *-schap/-ship* ($\chi^2=2.9$; *df*=1); *-iteit/-ity* and *-baar/-able* ($\chi^2=0.9$; *df*=1), but were significant between *-iteit/-ness* and *-achtig/-like* ($\chi^2=22.5$; *df*=1).

Table 18. Mean percentage correct scores for the affix pairs in the experiment

		Productivity			
		High		*Low*	
Translation Equivalence	*High*	-heid / -ness:	59	-baar / -able: -iteit / -ity:	69 81
	Low	-iteit / -ness: -achtig / -like:	21 32	-heid / -ity: -schap / -ship:	67 39

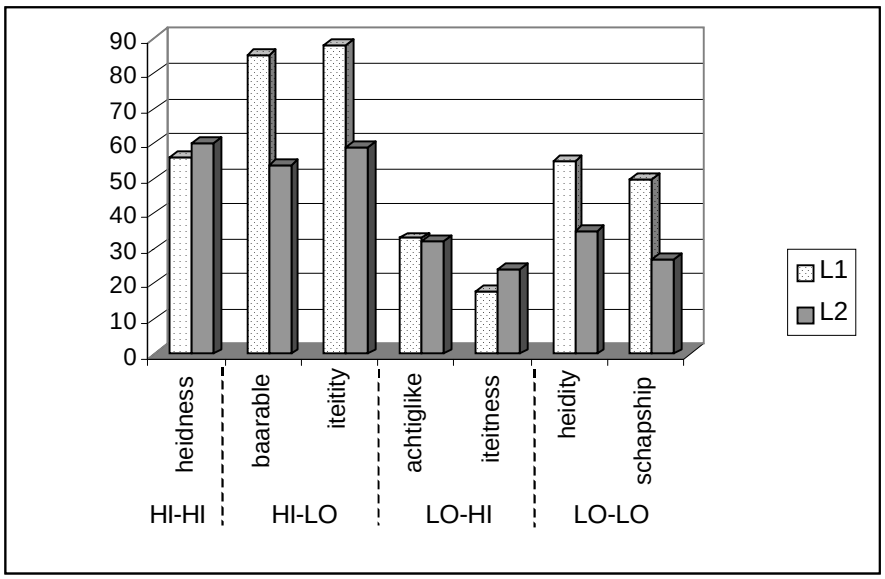

Figure 43. Mean scores for all affix pairs in the two tasks. At the bottom the respective levels have been provided for Translation equivalence and Productivity that the affix pair represents. The level of frequency for these data is "low".

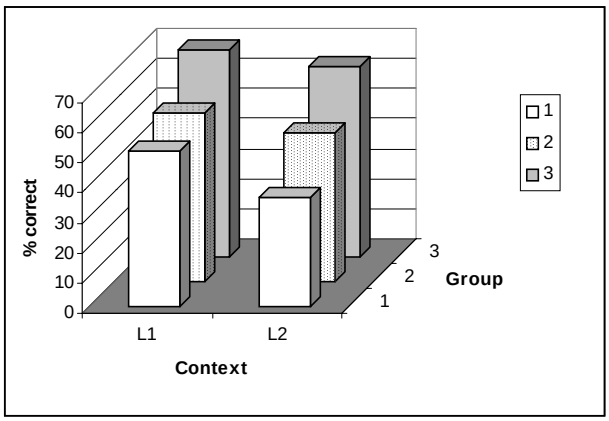

Figure 44. The overall differences with regard to the proportion of correct scores on the target affixes between the three groups participating in the experiment.

4.4.2.3.1 Groups

The difference between the groups (see Figure 44) turned out to be significant (F[2,113]=20.32; p<.01). Larger proportions of correct scores were found at higher levels of proficiency.

4.4.2.3.2 Context

The overall scores in the L1 context were higher than the scores in the L2 context. This difference turned out to be significant (F[1,113]=9.98; p=.002). Moreover, a significant interaction was found between context and group (F[2,113]=6.49; p=.002). The context effect was stronger at higher levels of L2 proficiency: in group 3 only a minor difference was found between the scores on the L1 task and the L2 task (see Figure 44).

4.4.2.3.3 Translation equivalence

The effect of translation equivalence turned out to be significant (F[1,113]=246.7; p<.001). That is, the proportion of correct scores increased with increasing degrees of translation equivalence. A significant interaction occurred between group and translation equivalence (F[2,113]=5,75; p=0.004). This interaction is graphically represented in Figure 45: the differences between the groups was larger at high level of translation equivalence, or, in other words, the difference in the effect of translation equivalence was strongest at higher levels of proficiency. No interaction was found between context and translation equivalence (F[1,113]=1.22; p=.27).

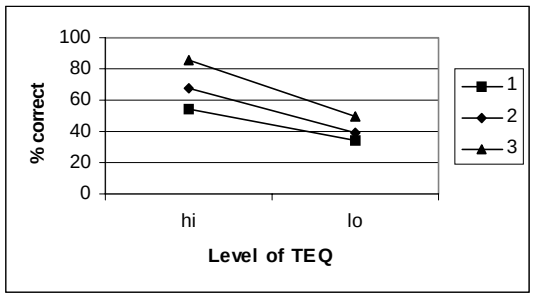

Figure 45. The interaction between the levels of translation equivalence and the groups.

4.4.2.3.4 Productivity

The effect of productivity was significant (F[1,113]=99.2; p=<.001). However, the largest proportion of correct scores was found at low levels of productivity. A significant interaction was found between group and productivity (F[2,114]=22.6; p<.001). The largest (negative) effects for productivity were found at lower levels of proficiency (see Figure 46). Further interaction was found for Context x Productivity (F[1,113]=79.72; p<.001): the negative effect of high productivity was stronger in the L1 context. The three-way interaction Group x Context x Productivity (F[2,113]=3.09; p=.049) showed that the strongest interaction between Context x Productivity was found at the highest level of proficiency. The interaction of Context x Productivity was very peculiar for Group 3 (see Figure 47).

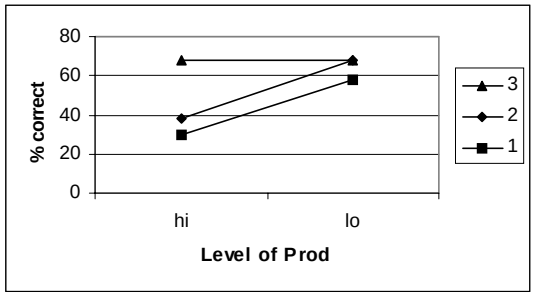

Figure 46. Interaction between group and productivity.

4.4.2.3.5 Higher order interactions

Several significant interactions were found between the main variables.

First, context interacted significantly with productivity (F[1,113]=79.72; p<0.001). In the L2 context, hardly any effect of productivity could be determined, while in the cross-linguistic context ("L1") a negative effect of productivity was

found: larger proportions of correct scores coincided with low levels of productivity. The three-way interaction between context, translation equivalence and productivity was also significant (F[1,113]=4.07; p=.046): the interaction between translation equivalence and productivity was strongest in the L2 context, and showed a reversed effect in the L1 context (see Figure 47).

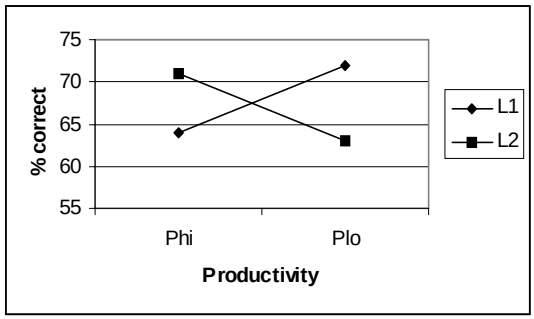

Figure 47. Interaction of context x productivity in group 3.

Second, a significant interaction was found between translation equivalence and productivity (F[1,113]=11.65; p=.001). The (negative) effect of productivity was strongest at the low level of translation equivalence (see Figure 48). In the L2 context, no productivity effect was found at the high level of translation equivalence.

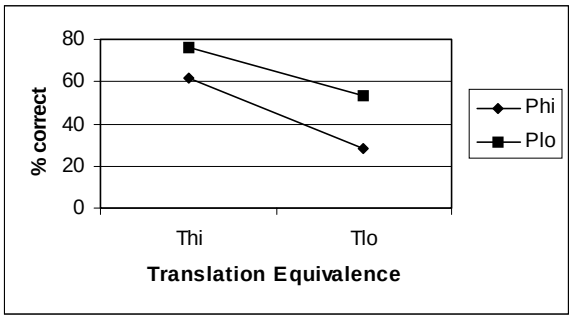

Figure 48. Interaction between translation equivalence and productivity.

4.4.2.3.6 Frequency

In spite of the attempt to restrict item selection to words with frequency values below 10, this appeared not to be possible for the less productive L2 affixes. It is for these affixes that somewhat higher word frequencies had to be selected (see *Figure 50*). Consequently, no absolute frequency values could be included into the

MANOVA. Yet, a strong correlation was found between absolute item frequency and the percentage of correct scores (r_{xy}=.81; p=<0.001, two-tailed; see Figure 49.

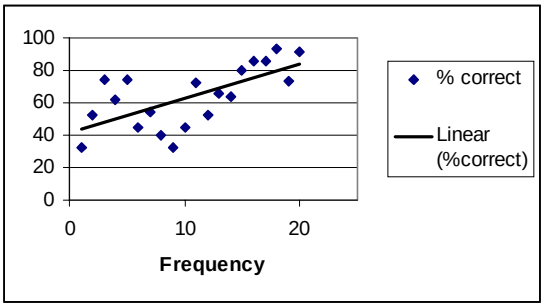

Figure 49. Correlation between the absolute frequency and the overall percentage of correct scores

Although absolute item frequency could not be included in the experiment, an alternative measure of frequency was included in the MANOVA applied to analyse this experiment. This (relative) measure of frequency was based on a division into degrees of frequency within each affix pair. For instance, at the high level of translation equivalence and the high level of productivity (HIHI in Figure 50), frequencies between 0 and 8 were categorised as "low" and frequencies between 10 and 13 were categorised as "high". For the HILO levels, however, that line had to be drawn between 1 and 2 (see Figure 50); value 1 was attributed to low and value 2 was attributed to high. Frequency thus categorised turned out to be significant (F[1,113]=83.01; p<0.001) and significant interactions were found with frequency and context (F[1,113]=10.54; .002), productivity (F[1,113]=46.83; p<0.001), and translation equivalence (F[1,113]=13.12; p<.001).

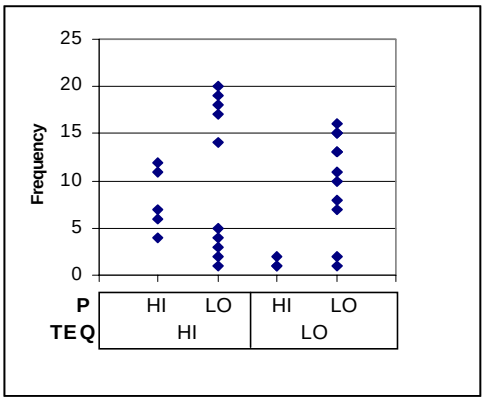

Figure 50. Distribution of levels of productivity (P) and translation equivalence (TEQ) over the item frequency of the target items.

In all cases, larger proportions of correct scores were found at the high level of frequency. The effect of frequency in the L2 context was stronger than the effect in the L1 context, at the low level of productivity, and at the low level of translation equivalence. The interaction between translation equivalence and productivity was strongest at the high level of frequency.

4.4.2.3.7 Syntactic category

The other dependent variable, syntactic category, is related to the score on the target affix. When the target affix was used, the syntactic category was always correct. In total, the correct syntactic category had been selected in 83 per cent of all scores (73, 90 and 91 per cent for groups 1, 2 and 3 respectively). Further analyses were run to investigate if a different affix was used of the same syntactic category, like a more productive affix or an affix which forms a better translation equivalent with the L1 affix. A cross-tabulation showed that for 68 per cent of all incorrect scores an alternative affix was used of the correct syntactic category. A MANOVA was run to investigate those cases where the targets affix was not used, but an alternative affix of the same syntactic category. The between-subject variable, Group, turned out to be significant at $p<0.05$: higher levels of proficiency more often filled in an alternative affix of the correct syntactic category ($F[2,5]=13.7$; $p=0.009$). The effect of translation equivalence, productivity and context were not significant at $p<0.05$. Significant interactions were found for translation equivalence x productivity ($p=0.033$; see Figure 51), translation equivalence x productivity x frequency ($p=0.013$) and group x translation equivalence x productivity x frequency ($p=0.017$). Finally, the scores of the affix pairs were considered (see Figure 52). The difference between the affix pairs was significant ($F[6,222]=11.46$; $p<0.001$). The largest proportions of correct syntactic category were found for the affix pairs -achtig/-like and *-iteit/-ness*. In the majority of cases, alternative affixes were used with higher values of translation equivalence, in spite of their lower productivity: *-achtig/-ish* and *-iteit/-ity* respectively.

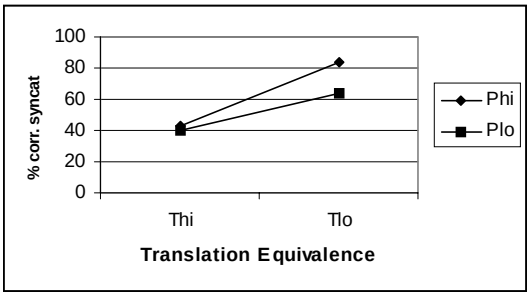

Figure 51. Interaction of translation equivalence and productivity for cases where affixes were used other than the target affix, but of the correct syntactic category.

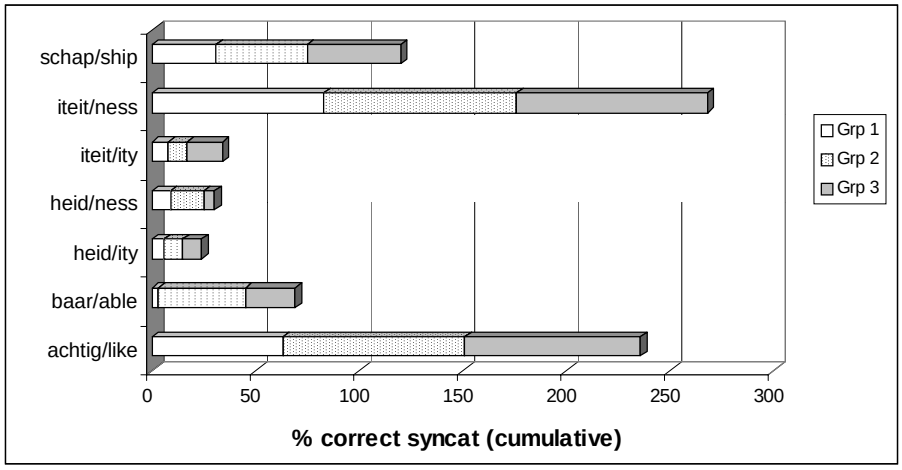

Figure 52. Use of an alternative affix of the correct syntactic category instead of the target affix. These percentages have been neutralised for frequency.

4.4.2.4 Discussion

The proportions of "correct" scores (i.e. use of the target affix) in this experiment were clearly affected by the main factors incorporated in the experiment: productivity, translation equivalence and frequency. These effects and their interactions will briefly be discussed in the light of the predictions stated earlier and the model presented in the previous chapters.

Apparently, the individual target items that had been selected to represent the affix pairs did not affect the results; the difference between the item sets was not significant. However, the order in which the items were presented did have a significant influence on the scores. This was a surprising finding in regard to the careful randomisation of the items. Since the orders had been equally distributed among all conditions for all subjects, the effect of the order will have been neutralised. This finding does indicate, however, that learners are very sensitive to affixes previously encountered. This is in line with the results of a pilot study carried out for the current experiment, where learners showed the tendency to use the first affix they were able to apply as the default type for the production of all morphologically complex words.

4.4.2.4.1 Frequency

The initial intention to exclude frequency as a variable in this experiment by exclusively selecting low-frequency target items did not succeed. Even the minor differences in COBUILD frequency incorporated in this experiment have clearly affected the resulting scores: higher proportions of correct scores were obtained at the high level of frequency. This observation corroborates the well-attested frequency effect: morphologically complex words that are highly frequent are not produced on the ba-

sis of type-familiarly, but have been given their own representation in the mental lexicon.

Besides the relative frequency effect incorporated in the analysis, the absolute frequency of the items may well have affected the results. The reason for this is that the absolute frequency appeared not to be equally distributed over the affix pairs or over the levels of productivity and translation equivalence.

4.4.2.4.2 Groups

An earlier experiment (see 4.2) showed that the type of naturally occurring groups used in these experiments did not only differ in terms of the level of L2 proficiency. Other between-subject variables like age and L1 experience may have affected the difference between the groups in the present experiment. Yet, L2 proficiency makes up important part of the differences between these groups. With this reservation, the variable Group in the present experiment will be regarded as an operationalisation of the level of L2 proficiency. The analysis has shown that the effect of the group was significant and that higher overall proportions of correct scores were found at the higher levels of this variable. This is in line with what can obviously be expected: the higher the level of proficiency of the learner, the higher the proportion of correct scores. However, the more interesting findings concerning this variable were found in its interaction with some of the other variables in this experiment. These will be reported in the sections concerned.

4.4.2.4.3 Context

The scores in the L1 context were generally higher than the ones in the L2 context. This finding is in agreement with what had been expected (see 4.4.2.2) and can be attributed to the nature of the task. In the L1 context, no interpretation of a sentence is involved, which decreases the chance of making errors. The interaction between context and group showed that learners at lower levels of proficiency have more problems at interpreting the English sentences than the learners at the higher levels of proficiency. At the highest level of proficiency, the difference between the scores in the two contexts has almost disappeared. This effect is not very surprising and in line with what had been expected.

4.4.2.4.4 Translation equivalence

The most important effect tested in this experiment was translation equivalence. The analyses show that the larger proportion of target affix use was found at the higher level of translation equivalence. This shows that in producing morphologically complex words in the L2, learners tend to rely on the equivalence relation between affix types in the L1 and the L2. This is in agreement with what had been predicted.

The interaction with the subject groups shows that the role of translation equivalence is strongest at high levels of proficiency. This, too, is in line with what had been expected. At higher levels of proficiency, learners have been exposed to English more and, therefore, have become more “aware” of the shared conceptual characteristics of affix pairs with a high degree of translation equivalence.

Contrary to what had been expected, no interaction was found between context and translation equivalence. It had been expected that translation equivalence would play a more prominent role in the translation task than in the L2 gap-filling task, as focus on the L1 affix might increase the activation of the equivalence with the L2 affix. Apparently, even in the L2 context learners are guided in their choice of the affix type by the degree of implicit equivalence to an L1 affix type.

4.4.2.4.5 Productivity

The effect of productivity was significant, but contrary to what had been expected, higher values of productivity lead to lower proportions of target affix use. The explanation for this contradictory finding must be sought in the interaction with other variables and in a methodological problem.

For the method employed in this experiment, a rigid division had to be made in finding affix pairs representing high and low levels of translation equivalence and productivity. However, the absolute values of translation equivalence and productivity were not always equally distributed across these levels. For instance, in the combination of high translation equivalence and high productivity, the respective values were 52/10. In the combination of high translation equivalence and low productivity, these values were 93/1 (for both affix pairs included). This means that the absolute values of translation equivalence may have affected the results in this category. Similarly, the average absolute value of translation equivalence was higher in the category combining low translation equivalence and high productivity than in the category where both variables were low (see Table 19). Because of the attribution of affix pairs to high/low levels, the higher absolute values of translation equivalence have affected the productivity scores. However, had productivity been a strong effect in itself, this problem would probably not have shown.

Table 19. Absolute values of translation equivalence and productivity across the levels.

		Productivity	
		High	*Low*
Translation Equivalence	*High*	52/10	93/1
	Low	18/14	27/1

The interference of translation equivalence was clearly less strong where several other factors interacted: at high levels of translation equivalence, in the L2 context, at high levels of proficiency and at low levels of frequency, either no effect of productivity was found or, especially where these factors interacted, productivity positively affected the proportion of correct scores. For instance, at high levels of L2 proficiency in the L2 context, productivity does have a clear positive effect on the proportion of correct scores. The effect of the context can easily be accounted for. In the L1 context, providing the L1 affix reinforces the interfering effect of the absolute values of translation equivalence. Obviously, in a pure L2 context L1 interference is less likely to occur. The effect of the Group is in agreement with what had been expected. At low levels of proficiency, the learners have not yet been sufficiently ex-

posed to the second language to reach a relative high resting activation of productive affixes. For these learners, there is no positive contribution to the production of productive types. In these groups, therefore, the interfering influence of the high translation equivalence values can indeed be expected to be strongest.

Finally, the high scores at the low level of productivity are further reinforced by the combination with the high absolute frequency of the low-productivity target items. Especially for the items of the *-ity* type higher absolute frequency values were found, and the data show a positive significant correlation between the absolute item frequency and the proportion of correct scores (see *Figure 49*).

4.4.2.4.6 Syntactic category

When a choice has to be made for the use of a particular affix type, it is obvious that learners in the vast majority of cases did manage to select a type referring to the appropriate syntactic category. In cases where an affix was used other than the target affix (the "incorrect scores"), 68 per cent of the alternative affix were of the appropriate syntactic category. For two affix pairs, *-iteit/-ness* and *-achtig/-like*, an alternative affix of the appropriate syntactic category was used in the majority of cases. Both of these affix types have a high level of productivity and a low level of translation equivalence. As predicted, in the majority of cases the alternative that was used instead of the target affixes was an affix with a higher value of translation equivalence that was less productive. This again emphasis the importance of Dutch-English translation equivalence for Dutch learners of English.

4.4.3 Conclusions about translation equivalence

For the experiment described in this section, the translation equivalence of several Dutch-English affix pairs was calculated in a corpus study. The translation equivalence of an affix pair was defined as the number of times a transparent morphologically complex L1 word comprising the L1 affix could actually be translated by an L2 word comprising the "equivalent" L2 affix, expressed in a percentage. Following the discussion in the previous chapter, it was claimed that translation equivalence would affect the production and comprehension of L2 words, due to interlingual co-activation of the concepts representing the affix types.

The second variable included in this experiment, productivity, was also calculated on the basis of corpus data. It was expected that the effect of productivity depended on the exposure to L2; L2 learners in the early stages of L2 acquisition have not yet acquired the productivity of all productive affix types in L2. Increased exposure to the L2 would lead to more highly activated lemma nodes of productive morphological types. Therefore, it had been expected that the role of productivity increase with increasing exposure to L2.

To test the effects of translation equivalence and productivity, three groups of Dutch L2 learners of English participated in two production tasks. The results of this experiment show that for these learners translation equivalence plays an important role in the production of morphologically complex words in the L2.

The effect of productivity was not in agreement with the expectations, due to the interference of the actual degrees of translation equivalence used in the experiment and the interference of the absolute degree of item frequency. However, in those conditions where productivity had been expected to be strongest, these interference effects were overcome. The fact that the strongest effect of productivity was found at higher levels of L2 proficiency confirms the idea of gradual acquisition of productivity. The finding that the strongest impact of productivity is found in the L2 context demonstrates that the effect of productivity is affected by L1 interference (i.e. translation equivalence).

With regard to the model presented in the previous chapters, it can be concluded that the effect of the L1 in the production of morphologically complex L2 words should not be underestimated. At all levels of proficiency, the production of L2 affix types is strongly dependent on the presence of a consistent equivalent affix in the L1. The facilitating effect of a consistent translation equivalent can be interpreted as evidence for the occurrence of overlapping semantic representations in the bilingual mental lexicon.

The frequency effect found in this experiment constitutes an influence of a different kind. Frequent morphologically complex items will have their own lexical representation, and no affix types are involved for the production of these words. The only affix characteristic that can be expected to affect the choice of these words is the syntactic category of the word.

4.5 Conclusion

In this chapter, several approaches have been used to investigate some implications of the model outlined in the previous chapters. The main emphasis has been to investigate the role of the first language in the type-familiar production and interpretation of morphologically complex words in the L2: how do several aspects of the first language contribute to the transparency of morphologically complex L2 words? To this effect three aspects have been investigated: the role of form-based vs. semantic similarity of L1 and L2 morphological types, the role of interlingual activation feedback, and relative contribution of translation equivalence and productivity to the correct type-familiar production of L2 words. In most of the experiments, some form of language development has been incorporated, to provide an insight into the process of the acquisition of L2 morphology over time.

In the first series of experiments it has been demonstrated that L2 learners are most proficient in the production and perception of L1 affix types that have similar representations in L1 and L2 in terms of form and are also semantically similar. Semantic similarity, i.e. an overlap between the properties of the L1 and L2 affix types for a particular L2 affix form, appears to be a prime condition for the acquisition and use of morphological types. If the overlap of conceptual properties is absent or is inconsistent, transparent production or interpretation of morphologically complex words in the L2 is hampered.

An attempt to measure the amount of interlingual activation in an experimental setting has failed. Probably the amount of activation is too small to be demonstrated

in the type of task administered and different techniques may have to be used. Moreover, the experiment yielded an inexplicable interaction between the random groups and the priming effect. This seems to indicate that the technique applied is not sufficiently reliable. More research has to be done to determine the interfering variable(s) that caused this effect. Another effect that was found in this experiment was that the reaction times on the less productive affix type *-able* was considerably faster than the more productive affix type *-ness*. The advantage of *-able* over *-ness* can be accounted for in terms of translation equivalence. In a third experiment, the degree of translation equivalence was determined on the basis of a corpus study. Analogous to the definition of productivity, translation equivalence was defined as the chance that a particular L1 affix form can be translated by using an equivalent L2 affix form. Since the starting point is that one particular form represents one particular type (except for homonyms), the consistent translation equivalence of an L1-L2 affix pair implies a high degree of overlap of the morphological types in L1 and L2. Translation equivalence is the quantification of the degree of similarity expressed in the first series of experiments. The translation equivalence study shows that a high degree of consistency of translation equivalence largely determines the extent to which L2 words are produced and interpreted type-familiarly. This study also clearly shows that L2 productivity is much less important for the L2 learner than the perceived transparency based on the L1, especially for beginning learners. An interfering factor in this experiment was the strong impact of item frequency: even small differences in (a low degree of) productivity appeared to have affected the results. The frequency effect especially interfered with the objective measurement of productivity, as low productivity inherently coincides with high frequency. The fact that in the priming experiment the affix type *-able* outperforms *-ness*, both in reaction time and in the number of errors, confirms the conclusion that L1-induced transparency (i.e. translation equivalence) takes precedence over productivity.

In the first series of experiments and in the translation equivalence experiment several between-subject factors had been included. Both experiments show that learners at higher levels of L2 learning are better at using L2 morphological types. However, it could not be consistently determined that this effect was due to the exposure to English (including the number of years of formal instruction in English) or to some age-related factor. The correlation that was found between a subject's performance in L1 morphology and in L2 morphology seems to indicate that the increased morphological skills are (partly) due to experience and development in L1. The translation equivalence experiment shows that learners at higher levels of proficiency are increasingly affected by L2 productivity, especially in contexts where the L1 does not interfere. The same holds for translation equivalence. Both of these factors are gradually acquired as a result of increased exposure to the L2.

It should be noted that the production and interpretation of morphologically complex L2 words in the experiments described here are examples of forced production and the results cannot be generalised to everyday "spontaneous" discourse. However, if no effect of type-familiarity had been found in these experiments, these effect would not be likely to occur in spontaneous speech either, as the experimental settings in most of the experiments achieve a maximum of attention on morphology.

Once type-familiar interpretation of morphologically complex words in the L2 has been established, further studies can be devised to investigate this phenomenon in more natural contexts.

In conclusion, it can be argued that the transparent interpretation and production of morphologically complex words in L2 is strongly dependent on the extent to which the L1 and L2 affixes are similar in form and consistently represent overlapping semantic representations. Through psychotransparency, the learner's first language is of primary importance in the acquisition and use of morphological types. It is only at later stages of L2 acquisition that L2 productivity starts to play a significant role in this.