Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategy.
Poelmans, P.

Citation for published version (APA):
Chapter 6  Conclusions and suggestions for further research

6.1. Recapitulation of research questions and main findings

The study described in this thesis was mainly set up to compare the effect on listening comprehension of a training of lower-order skills with a training of higher-order strategies. Closely linked to this topic is the claim of Segalowitz and Segalowitz (1993) arguing that automatisation of the recognition of (most of) the words spoken is conditional for successful listening comprehension in a second language. The study described in this thesis can therefore also (partly) be seen as an empirical study to investigate the validity of that claim.

A first step that had to be taken in order to find an answer to the main question described above was the development, in Pilot Study I (Chapter 3), of a criterion to distinguish between automatised and non-automatised processes. The criterion that we developed was based on the assumption that the process of word recognition is fully automatised in one’s mother tongue but not in a second language. It was discussed in Chapter 2 and Chapter 3 that intensive training could cause the word recognition processes to become more automatised. In that case the question is whether it is possible for the L2 word recognition process to become fully automated (near-native), or whether there will always be a distinct difference in degree of automatisation between L1 and L2 speakers. To answer the research question an auditory lexical decision experiment was set up. The experiment in which 25 native speakers and 25 non-native speakers of Dutch participated, included despite the distinction between words and nonwords, also overarticulated versus underarticulated items. As was described in Chapter 3, the main difference between overarticulated and underarticulated speech is the degree of carefulness in the pronunciation whereby overarticulated items are pronounced very carefully whereas underarticulated speech can be defined as sloppy speech. In every condition (i.e., overarticulated words, overarticulated nonwords, underarticulated words, and underarticulated nonwords) there were 1, 2, and 3-syllable items, for example *huis* ‘house’, *ogant*, *omgeving* ‘environment’. First we tried to make a distinction between the (automatised) processes of the native speakers and the (non-automatised) processes of the non-native speakers, by using the measure of automatisation that Segalowitz and Segalowitz (1993)
suggest, namely the Coefficient of variability and the correlation between this coefficient and the mean RTs. The results of using this measure showed only a perfect distinction between native and non-native speakers when the performances of a number of participants were excluded from further analyses. Since we preferred using all available data and since we think that besides speed, accuracy is also an important indicator of automatisation we decided to develop our own criterion to measure automatisation. Discriminant analyses based on the results of the lexical decision test showed that it is possible to make an almost perfect distinction between the performances of native and non-native speakers in the recognition of one and two-syllable overarticulated nonwords. Therefore we could conclude that the criterion to distinguish between automatised (L1) and non-automatised (L2) processes can be found in the speed and accuracy of participants' reactions on one and two-syllable nonwords. The criterion found was used in the further experiments of this thesis.

A second step in the way to answer the main research question was the investigation, in Pilot Study II (described in Chapter 4), of the relation between language knowledge, status of the word recognition process (in terms of speed and accuracy) and general listening comprehension skills. More concretely, the question was whether it is possible to categorise language learners into fixed categories based on the performances on tests assessing the language knowledge and the skills described above. We distinguished three possible categories. Participants in category 1 were assumed to be learners with a good knowledge of the language that are able to use this knowledge under time pressure, they were assumed to have good listening comprehension skill. Participants in category 3 were assumed to be learners with poor results on the knowledge tests, by definition, automatisation of knowledge is therefore not possible, we predicted that they had a poor listening comprehension skill. Finally, participants in category 2 were assumed to be learners with sufficient language knowledge that lack the aural word recognition skills necessary to successfully use the knowledge under time pressure, they were therefore predicted to have a poor listening comprehension skill. In addition to the investigation of the relation between knowledge and skills, Pilot Study II also investigated the complementary value of on-line tests, which are defined as time-critical, to off-line tests, in the compilation of a detailed language proficiency profile of a language learner. A test-battery that included knowledge tests, a Listening comprehension test, an on-line Lexical decision test and a memory test, was administered. The performances on these tests of the 20 second language learners that participated in this pilot study indicated that it is not simple to categorize language learners into fixed categories. The relation between language knowledge, word recognition and listening comprehension was not
as straightforward as we expected; it was necessary to determine sub-categories in order to categorize the participating L2 learners. Six of the learners had sufficient language knowledge and the status of their word recognition process did also meet the criterion we determined, but they had a bad listening comprehension skill while four students also had enough language knowledge but despite that, the status of their word recognition process was poor while their listening comprehension skill did meet the criterion. The necessity to use sub-categories, however, indicated the usefulness of online tests for compiling language profiles; without the use of these tests wrong conclusions can easily be drawn regarding the profile of language learners. Without the use of the Auditory Lexical Decision test, for example, one could easily draw the wrong conclusion that the participants that did not pass the (offline) Listening Comprehension test were poor second language listeners. The results of the (online) Lexical decision test, however, made clear that these participants that performed poorly on the general listening test, had good performances on the online decision test. In other words, reaction times and percentages correct on the online tasks made it possible to make rather subtle differentiations between the language skills of L2-learners resulting in a detailed language proficiency profile. Furthermore, the results on the auditory lexical decision test of this second pilot study were in line with the results of the test used in the first pilot study. The parallel in the results of Discriminant analyses between the two pilot studies strengthens the validity of the criterion that was the result of Pilot Study I (Chapter 3).

The information gained in the two pilot studies described above was used in the training study (described in Chapter 5). It was already stated at the beginning of this chapter that the training study was set-up to investigate the differential effect of training lower-order processing skills versus higher order comprehension strategies. The study was set-up in three stages: Test stage 1 – Training – Test stage 2. During the test stages, knowledge tests (Grammar and Vocabulary) to establish participants' knowledge of Dutch were administered. There were also tests included (e.g. memory tests) the results of which were expected to indicate a relation between these tests and general listening comprehension skills. The most important tests, however, were the tests that focused on the listening comprehension process itself (the listening comprehension test) and its component processes: word recognition (Lexical Decision test) and sentence processing (Sentence Verification test). In this study, two experimental groups were compiled, a so-called Recognition group, which received a training in the processing of lower-order skills like word recognition, and a so-called Comprehension group, which received a training in higher-order comprehension strategies. A Control group was also included the participants of which were excluded
from the training; they participated, however, in the two test stages. Since at Test stage 1, no significant differences were found between the groups, it was expected that a conclusion about the training effect could be drawn based on the results obtained in Test stage 2.

In the training study, the notion of automatisation and its relation to listening comprehension was highlighted. The link for this study between automatisation of the word recognition process and listening comprehension lies in the claim of Segalowitz and Segalowitz (1993), which was also stated above. We reformulated their claim applied to our main research question in two hypotheses. The first hypothesis stated that the students that participated in the Recognition group will perform better on the post-training general listening comprehension test than would the participants in the Comprehension group. The second hypothesis stated that, even if no positive evidence could be found to support the first hypothesis, a significant difference will be found in the performances on the post-training Sentence Verification test or at least in the performance on the post-training Lexical Decision test since these tests measure processes that form the focus of the method of the Recognition group. However, the study did not show any of these effects. The results of the study did not give a clear answer to the main research question: no significant differences were found on the listening comprehension test between the two experimental groups. The data did also give no clear additional information about the relation of individual characteristics of the learners (e.g., the memory span) and listening comprehension.

Based on the results obtained in the present study we can not conclude that training lower-order skills is more beneficial for second language learners than training listening comprehension in the more traditional (comprehension) way. However, neither do the results allow the conclusion that the traditional method is better. It is reiterated here that the training study should not be regarded as a comparison of two methods competing for educational relevance but as a comparison of training effects of two theoretically relevant component processes. As we did not find support for our hypotheses, there is even less reason to draw conclusions from this study for L2 instruction. As we stated in chapter 5, there are theoretical reasons to claim that listening comprehension can best be acquired through a combination of training in processing information at lower levels of information (up to the sentence level) and a training in applying strategies in dealing with information at higher levels of information (beyond the sentence level). However, as long as solid empirical evidence for such a claim is lacking, it is prudent for researchers to refrain from giving advice to educationalists. Since the results on the Staatsexamen
NT2 (as was described in Chapter 1) do indicate that the traditional training method is not optimal, finding a more satisfying method is still necessary.

6.2. Suggestions for application of the findings and for further research

The overall results of the present study do not allow us to recommend a change in the training of general listening comprehension skills towards training in lower-order skills such as word recognition. However, we would like to make a few recommendations based on the findings described in this thesis.

A first recommendation concerns the use of on-line tests in compiling a detailed proficiency profile of a language learner. The usefulness, or even the necessity, of using these time-critical tests providing information about the underlying processes of a skill, was discussed in Chapter 4. It was shown that when the results on the on-line tests were not included in the assessment of the learners’ language proficiency, erroneous conclusions could be drawn concerning the skill tested. We would like to suggest that the use of on-line tests might uncover some of the weaknesses of a language learner’s language command; once this is known, the language teacher can develop a specific training programme to foster the learner’s language skill. In the study described in this thesis, the listening comprehension skill was the target of the investigation; we, therefore, mainly used auditory tests. Future research could concentrate on, for example, the visual mode or on production. If the findings of such research would be in line with the results presented here, it would be possible to develop a test battery and a method of data analysis and categorisation such that a general (language proficiency) profile of a language learner could be formed.

Concerning the development of the automatisation theory, we think that our results contribute to the idea that automatised and controlled processes are not a dichotomy but that there is a continuum between the two extremes. The results of the LDA-analyses described both in Chapter 3 and in Chapter 4 show a continuum between the L1 and L2-speakers’ performances (expressed by mean RT and percent correct) such that performance of the L1-speakers were interpreted as automatised and performance of the L2-speakers as more controlled. The distance between native speakers’ and non-native speakers’ performance supports the assumption that the process of spoken word recognition is automatised in one’s mother tongue but not in a second language. We therefore think that using LDA-analyses based on the mean results (RT and percent correct) of a
lexical decision test of L2-speakers and L1-speakers produces an index of the status of the spoken word recognition process.

In the discussion and conclusion section of Chapter 5 (section 5.4.4), the chapter in which the training study was described, a few reasons for the lack of clear training effects were given. It was said that the most likely reason why neither of the two training methods had a positive impact on listening comprehension was lack of overlap between the materials used in the training and the materials used in the tests. This was described as a lack of transfer appropriate learning. It was stated that we may have been too optimistic in assuming that the similarity between the cognitive processes at the time of learning and processes at the time of testing would be sufficient to allow successful transfer of training. There was, of course, some overlap, as some of the words used in the training also occurred in the texts used in the Listening Comprehension test. This overlap was, however, not at the focus of our attention. In further research this focus could be established by manipulating the overlap between training materials and test materials. Manipulating the overlap in a way that the amount of overlap is in focus might result in clearer results concerning the effect of training. If this would be the case, the results would subscribe the transfer appropriate learning theory.

In the discussion section of Chapter 5, the remark was also made that it is rather impossible to control the actual amount of input that participants receive. The participants of the experimental groups in the present study received a lot of input during the intensive training and all participants followed a Dutch course in which they received a lot of input. The uncontrollable aspect concerning the input is the amount of input the participants received in contacts outside the classroom and the test situation in which they were confronted with Dutch (active or passive). A possibility to control the amount of input to a larger degree and that way get more clear results is by conducting the training in another language environment.

One last suggestion concerns the fact that the results of the training study do not indicate the importance of individual characteristics of language learners (i.e. memory span) in relation to listening comprehension skills. This might be due to the fact that the tests we used were either completely in Dutch (Listening span test) or they followed the grammatical and phonological rules of Dutch (Serial recognition test). It might be useful, in further, research to use also memory tests in the participants L1, the results on these tests might give insight in the relation between listening comprehension skills and the individuals’ memory span. Using L1-memory tests and comparing the results on these tests with results on L2-memory tests, might give information about the language dependency of the memory
capacity, a very interesting research topic the results of which might give insight in the importance of memory span in language learning.