A dynamic approach to the development of lexicon and syntax in a second language
Chan, Belinda

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 8

Conclusion
The aim of this dissertation was to apply the dynamic usage-based approach to investigating second language development. We investigated what factors may play a role in DST by observing whether beginner learners in Taiwan can enhance their vocabulary knowledge when given extra natural exposure to input (enhanced input). We also investigated how the language system and its interrelated components are shaped over time by observing the development of writing and speaking in terms of lexical and syntactical complexity. In order to address the above questions, four related studies were carried out.

8.1 The effects of enhanced input on learning collocations and associations

In the first study, we investigated the effects of enhanced input on beginner learners’ contextual vocabulary knowledge, looking specifically at collocations and associations. Moreover, as our experiment took the experimental group as its own control group at different moments in time, the effect of the sequence of the input conditions on learning collocations and associations was also investigated. Two homogeneous groups of beginner learners of English in Taiwan were assessed at three moments in time on four dimensions of vocabulary knowledge: receptive collocations, productive collocations, receptive associations, and productive associations. Group A received five hours of extra enhanced English input per week in the first period of four months but did not receive any extra enhanced English input in the second period of four months. Group B received the inverse input conditions. The results showed that the natural exposure to English input did not necessarily lead to learning of contextual vocabulary knowledge for beginner learners of English. However, the sequence of input conditions had an effect on learning contextual vocabulary knowledge, in the sense that group A showed a decrease from low to high input whilst Group B showed an increase from low to high input. Group B happened to establish the link between the word form and meaning of the target items in the first four months (period of low input), and thus the enhanced input, presented after the link between word form and meaning was established, helped learners to learn the contextual vocabulary knowledge.

From a usage-based perspective, input plays an important role in the development of a second language. As input turns into intake when input can be comprehended by learners, the knowledge of linking word form and meaning, if presented in advance, helps learners to comprehend the input to the meaning level. The comprehension of input at the meaning level thus improves the degree of turning input into intake. This explains why what seems to be a more important factor in SLD in this study is the sequence of input conditions rather than the input itself. The sequence of presenting the input should start with form and meaning and then move to the contextual knowledge such as collocations and associations. Moreover, input turns into intake when there is an sufficient number of repetitions. Our
study found that extra five-hour sessions of enhanced input may not offer sufficient repetitions to enhance the learning of contextual vocabulary knowledge. This may be due to the fact that the acquisition of contextual vocabulary knowledge needs not only more repetitions but also more engagement with the input.

8.2 The effects of enhanced input on the difficulty of word use in writing

The first study has demonstrated how input may shape learners’ use of contextual vocabulary knowledge and how input may be noticed through providing different vocabulary knowledge at different times. In the second study, we intended to take a closer look at the processes of how input may shape learners’ habit of using words of differing difficulty levels, which is determined by the number of words in different frequency bands. We investigated whether beginner learners of English use more difficult words, operationalized as vocabulary size, in their writings when they have more input and whether they use less difficult words in their writings when they have less input. We followed four beginner learners of English, who received different degrees of enhanced input over five months. We measured the difficulty of word use in 56 writings and quantified the developmental patterns of four learners. The results showed that learners of English improve their use of lexical difficulty in writing when they receive more exposure from input outside the classroom. When there was more input, there are more difficult words used in writing productions. It confirms the characteristic of a dynamic usage-based approach that language development is iterative. What learners have been exposed to (input) has become their language use (writing); their language use then becomes further input into the language system.

DST looks at the process of language development. With such dense longitudinal data, the application of the mathematical model in explaining the developmental trajectory seems to give more explanatory power than a descriptive analysis. However, the selection of the mathematical model is fundamentally important: it should be able to explain the developmental trajectory in the sense that it can capture all the characteristics of this individual. This study applied one mathematical model (a logistic model) to quantify the developmental patterns with three parameters: initial value, learning rate, and carrying capacity. We evaluated whether the logistic model can adequately describe the learning trajectories of four learners and found that it successfully described the vocabulary development of three participants in this study. The only exception was the participant whose developmental pattern was more complex than the logistic model. Therefore, the logistic model was not able to describe the development of that particular participant.

This result indicates that other possible variables in the logistic model (such as the degree to which learners are able to absorb the vocabulary from the exposure) should also be considered and also that vocabulary knowledge may be lost.
From a dynamic usage-based perspective, there are many more factors involved in determining the pattern of vocabulary learning (i.e. hours of input or degree of absorbing input), and thus learners develop their own particular way of constructing their language. Each particular way aggravates the difficulty of building up a general mathematical model which can describe the particular developmental pattern of each individual. If every individual has his/her own specific pattern, it means that the model must be able to describe many kinds of developmental patterns. A model which can describe many kinds of developmental patterns must include as many crucial variables as possible. However, adding the number of variables (a more sophisticated model) makes it more difficult to find the best set of variables due to the fact that more and more local minima emerge in root-mean-square-deviations (RMSD) calculation. Therefore, in order to reach the global minimum of RMSD calculation to obtain the best set of variables of the model, in principle, one needs to start fitting procedures with different initial values as many as possible. But, in practice, because of finite trials and iterations, it is very challenging to determine the best set of variables when a such complex nonlinear model is introduced.

The result in the second study also shows the difficulty of interpreting the longitudinal study. Although four participants improved the difficulty of lexical use in writing with different degrees of enhanced input, they did not show similar developmental pattern. That is, the product (result) of the answer is the same for these four subjects, but the processes of the development differ—only Grace and Gloria (identical twins) showed very similar developmental patterns in lexical difficulty.

8.3 Lexical development in writing and speaking

The first two studies have informed us that input indeed impacts on learners’ use of vocabulary: contextual vocabulary knowledge and the difficulty of word use. In the second study, we have demonstrated how the logistic model can be applied to explain the language development and elucidate the difficulty of sorting out a general mathematical model to explain the developmental trajectory for each individual, as every person has her/his own developmental path. In the third study, we intended to find out how the embedded subsystems (writing and speaking) are interrelated in terms of lexicon by investigating whether the development of writing is more advanced than that of speaking; and how these two subsystems correlate over time in lexical diversity and lexical difficulty. Two identical twins produced 100 writings and 100 recordings over eight months. We looked at the theoretical models of writing and speaking processes and found automaticity to be the major difference between the two tasks in our study. The time allowed to search for words is evidently less in the monologue speaking task. The results demonstrated that writing showed greater variety of word use than speaking but did not show greater difficulty of word use than speaking. The dynamic correlations of the two types of lexical complexity were described through a mathematical model (the hidden
Markov model), but the directions of the relations between writing and speaking were not similar among the identical twins.

From a dynamic systems theory perspective, the subsystems (writing and speaking) are interconnected. In this study, we found this interconnectedness changed over time. Writing and speaking may compete in one learning stage but may support each other in another learning stage. Moreover, from the dynamic usage-based approach, each individual has his/her own way of developing the language. In this study, we had a very interesting finding given the fact that the two informants, who were identical twins, did not demonstrate the same directions of correlations over time. We argued from this finding that understanding individual development is as important as understanding group performance and mean tendencies.

The application of the mathematical model (the hidden Markov model) on investigating the dynamics of the subsystems seems to quantify the language development well enough for every individual as this model does not require resembling trajectories between the mathematical model and the data. The Hidden Markov model clearly determines the three learning stages and provides useful information on every learning stage. However, there is a need to further investigate the number of stages required for the most efficient implementation of this model. The third study, which was based on the second study, took three learning stages as the ideal number to reveal the hidden learning stages as the number of learning stages is what has been assumed in the logistic model applied in the second study. If possible, it is important to test the number of the learning stages needed to reveal language development. But of course, this is again related to the individual differences found amongst participants.

### 8.4 Syntactical development of writing and speaking

In the third study, we explored how the embedded subsystems (writing and speaking) are interrelated over time in terms of lexicon, and successfully applied the hidden Markov model to investigate the dynamics of the subsystems at three learning stages. Although lexicon and syntax are two inseparable subsystems, we can only understand the dynamics of the language by first studying each subsystem (lexicon, syntax) in isolation and then studying how these subsystems are interrelated.

The fourth study was a follow-up study of the third study: it investigated how the embedded subsystems (writing and speaking) are interrelated over time in terms of syntax with the same data of the identical twins. We investigated whether the development of speaking is larger than that of writing in three types of syntactical complexity (mean length of T-unit, dependent clause per T-unit, and coordinate phrase per T-unit) and how the correlations of writing and speaking change over time. The results demonstrated that speaking showed greater syntactical com-
plexity than writing. The disadvantage of automaticity in speaking, as found for lexical diversity, did not seem to exist when looking at syntactical complexity. The dynamic correlations of the syntactical complexity were again described through the mathematical model (the hidden Markov model) and were again found not to follow similar directions between the identical twins. Therefore, we conclude that syntactical complexity is larger in speaking, and the correlation between writing and speaking changes over time, although the direction of the correlation may still remain unknown as two identical twins did not show one common direction.

On the basis of these findings, we argued that these measures of syntactical complexity indeed showed the complex nature of the syntax, in particular in the longer sentence units. However, a more complex syntax does not always indicate better quality productions. For instance, a learner may be able to produce a short precise sentence to reveal one idea. In this case, longer sentences do not necessarily represent higher quality or development in the language. Therefore, it might be interesting to observe how effective these learners can present their idea with their lexicon and syntax by investigating, for example, number of words per idea.

8.5 Contribution and future research

This study has shown that a DST approach is useful in analyzing different types of SLD data. Various mathematical approaches from the DST Toolbox can be used to analyse the data discussed, in particular the dense longitudinal data of our informants. The data also elucidates the interconnectedness of the different parts of the language system as they develop over time.

The present study is only the beginning of what could become an entire subfield of Applied Linguistic research: how the language system and its interrelated components develop over time, and what factors play a role from a DST perspective. As this study shows, dense longitudinal data allows us to model and analyse L2 development over time. The models and simulations are as yet rather crude. SLD is a very complex process with many factors on the social, psychological and contextual level. Only by building rational models in which our knowledge on what factors may play a role in development is included, will we be able to enhance our understanding of SLD. The models will allow us to combine factors that in real life are in fact out of reach. Ultimately, the models and their predictions will have to be tested on real data. Therefore, the need for suitable data continues to be one of the main issues in our study of language as a complex system and language development as a complex process.