Proposal for Research Traineeships 2015

1. Title of the Project

The interaction of words and constructions in artificial and natural language learning

2. Coordinators

Dr. Afra Alishahi (DCI)

Dr. Maria Mos (DCI)

Prof. Ad Backus (DCU)

3. Project Summary

Languages are complex systems and learning one consists of many different aspects. When learning a new language from scratch, humans must not only figure out what each word refers to, but they also have to discover the common structure of a sentence in the new language. As the learners receive more and more information, knowing each of these aspects will facilitate the acquisition of the other. For example, if a learner of English has already learned the meaning of words boy, ball and hit, upon hearing the sentence the boy hit the ball while watching the hitting event, they can infer that the word for the "doer" (or Agent) of an action usually appears before the word for the action, and the word for the thing on which the action is performed (or Patient) follows the action word.

On the other hand, once the learner discovers the "Agent Action Patient" (or "Subject Verb Object") order, it would be much easier to deduce the meanings of novel words which appear in this familiar construction: hearing the glip nit the blick while watching a duck pulling a rabbit, the learner can make an informed guess that glip must mean duck and blick must mean rabbit. This hypothesis is known as Syntactic Bootstrapping.

In this project, we will look at the interaction between acquiring word learning and structure learning in two ways. The first project uses the artificial language learning paradigm: experiments with rigorously controlled input in the form of a limited set of stimuli. The second project takes the approach of early foreign language learning.

Project 1:

Although there is ample evidence to suggest that humans are sensitive to cues from sentential context when learning new words, the interaction between learning words and learning grammatical constructions has not been carefully studied. In this project, we will design and conduct a series of experiments to investigate various aspects of this interaction, including:

- The characteristics of the input (e.g., frequency of certain words or constructions) which can facilitate learning the language;

- The properties of the words (e.g., nouns versus verbs) which may show different developmental patterns;

- The typology of the language (e.g., basic word order) that might affect the ease of learning.

The prospective student assistent will be actively involved in designing, executing and analyzing experiments to test the questions posed above. This means that, together, we will construct artificial learning tasks, including instructions, stimuli etc, then collect data regarding the acquisition of word meanings and structure knowledge, and analyze these data statistically. In an earlier project, material (visual and auditory stimuli, e-prime set-up etc.) was developed that may serve as a starting point for the project.

Project 2

Project 2 retains the focus on statistical learning and extends it to one of the domains experiments which artificial languages typically try to model: the learning of a foreign language from scratch. A group of native speakers of Dutch will participate in an experiment in which they learn either Russian or Chinese. In order to enable a comparison with Project 1, the learning conditions will be defined such that they are as similar as possible to how learners in that project learn the artificial language. In this study, we test the hypothesis that learning a 'real' foreign language is easier than an artificial language. First, we expect that there are features in natural languages that make them easier to learn than artificial languages. In evolutionary linguistics, the idea is taking hold that universal typological tendencies reflect learnability biases, and that deviant features are at an evolutionary disadvantage, and therefore remain rare or disappear entirely. Typical candidates for such features are the presence of a difference between subject and object, between nouns and verbs and other features that are often portrayed as universal or near-universal. However, there are no widely accepted theories about the learnability of linguistic features, and presumably must await results from a new form of typology that looks beyond the traditional focus on grammatical structure, to also take on board lexical ineffability, discourse preferences, etc. It might be that natural languages such as Russian and Chinese will have these features while our artificial language may lack them; hence, participants learning these languages should perform better than participants who learn the artificial language, no matter how carefully we design it. Moreover, we test the hypothesis that participants learning Russian should have more support from knowing Dutch than the learners of Chinese, given the fact that Russian is an Indo-European language and Chinese is not.

Overall goals and deliverables

The project has two overall goals. The first is to study the acquisition of word meaning and constructional patterns, and the degree to which they interact, when a learner starts learning a language from scratch. The whole team will jointly produce two working papers detailing the results of each project.

The second goal is to help improve external and ecological validity of artificial learning experiments. We will not really compare the learning of artificial and natural languages as such, since that would require a much more elaborate design. Instead, we wish to explore to what extent typical artificial language learning paradigms such as ours really model natural foreign language learning, in order to learn some lessons about improving the design of artificial languages in future simulation studies. Our expectation is that there are things in natural languages that make them easier to learn than artificial languages, such as discourse structure and the way pronouns are used. Knowing more about this would lead to implications for how to better design artificial learning experiments. We will report on the findings in a third paper, again as a joint production of all team members.

4. Project timeline

 List of planned activities (possibly with milestones and intermediary results).

Month	Project 1	Project 2	Milestone
1	Literature study		
2	Design of experiment 1	Literature study	
3	Collect data		
4	Analyze data	Design experiment	
5	Write report		
6	Design of experiment 2		
7	Collect data	Collect data	
8	Analyze data	Analyze data	Finish experiments
9	Write report	Write report	
10		Compare the studies	First working paper
11		Write report	
12			Second working paper

5. Research Trainee Profile

The prospective trainees will have some affinity with the field of language acquisition, and preferably have basic knowledge of the usage-based approach in linguistics. Also, they should have had some basic methodological training in experiment design and statistics.

The candidates must send their resume, a copy of their trascripts and a one-page motivation letter by email to the project coordinators ({A.Alishahi, Maria.Mos, A.M.Backus}@uvt.nl).