

Master's thesis review

Charles University, Faculty of Mathematics and Physics

Thesis author Daragh Meehan

Thesis title Learning Multiple Languages Through Parallel Audiovisual Input

Year 2024

Study programme Computer Science

Study branch Language Technologies and Computational Linguistics

Reviewer Josef Jon

Department Institute of Formal and Applied Linguistics

Review:

The thesis explores the use of audiovisual inputs in learning foreign languages, especially more languages at once. There are two main outcomes: a study of the effectiveness of learning multiple languages from short captioned and dubbed videos and software that allows the extraction of additional learning materials (flashcards, conversations) from such videos.

The introduction is well-structured and provides an overview of the historical context and current gaps in the field of language learning. The author argues for the relevance of using familiar content in L2 learning and provides a strong motivation for the following work. The use of historical references, such as bilingual texts from ancient civilizations, adds depth to the argument and demonstrates the long-standing tradition of parallel language learning. The author proposes a list of research questions that are answered by the thesis. Referencing these questions in the following section binds the whole thesis and links the experiments with the developed software. While similar work on the topic of audiovisual learning resources is quite extensive, the idea of learning multiple foreign languages at once is novel. Overall, this introduction identifies the research gap and proposes a clear set of thesis objectives.

In the second section, the concepts are grounded in psychological, cognitive, and second language acquisition theories, giving more validity to the following results and helping to interpret them. The background is very well-researched, looking at L2 learning from multiple angles and covering related work stretching from the initial publications on the topic to current (2023) research. The author bases the requirements for the software developed as a part of the thesis on this background research. One shortcoming of this section is that the statistical devices used in the study (e.g. Cronbach's alpha, Levene's test, Dunn's test, Bonferroni correction, etc.) are not introduced. A short description of the methods would make the thesis more self-contained.

The third section presents the design and both the qualitative and the quantitative results of the study carried out by the author. The goal is to assess the effect of watching a captioned video

in multiple languages (one episode of Peppa Pig TV show) on language learning. The participants watch the video first in English (their native language) and then in Italian and Dutch (either one-by-one or interleaved shorter segments) with captions in the respective language. Newly learned vocabulary is measured by form recognition and meaning recall test.

A larger sample size would give more validity to the experimental results, but the methodology is sound and I believe the extent of the study is more than sufficient as a proof-of-concept for a Master's thesis. The difference between the control group and both Segment/Whole groups is large enough to suggest learning effects even for such a small number of participants. On the other hand, both experimental groups had better results even on the distractors compared to the control group, suggesting better prior knowledge of Italian and Dutch in these groups. It is hard to tell if the differences between the groups stem from the individual capabilities of the participants, or from the learning method used for the group. The author even provided a significance analysis of the results. This issue is well-addressed by the use of covariate models and controlling for the previous knowledge. The author addresses all the issues with the sample size and inter-group differences in Section 3.6.1. The author is aware of the limitations of their experiments and the design of other possible experiments that were out of the scope of the thesis is described as well. A significance analysis of the results is also provided. Overall, I believe that given the constrained (zero) resources for the participants, the author did an excellent job in designing the study, analyzing the responses, and providing as many insights as possible.

The fourth section presents an overview of the design and the functionalities of the program developed during the thesis. The program processes video files with multiple audio and subtitle tracks and creates a database of resources suitable for further learning. These resources include interactive parallel texts, multimedia flashcards, and audio segments of dialogues. The software also integrates third-party dictionaries and machine translation APIs. The program actually works (the installation difficulty is reasonable and the software is multi-platform) and it is user-friendly, I believe it can be used in practice to learn multiple languages at once. The architecture is based on the Model-View-Controller design pattern, which is appropriate for this use. The code is readable and commented well. One of the more challenging problems is the alignment of the subtitles in different languages and the creation of reasonably long segments. The rule-based algorithm is explained in the text of the thesis. One small issue is the relationship between the experimental section and the software itself – the study is focused on different, although related, types of language learning, and the software itself is not evaluated by it.

Overall, I have really enjoyed reading the thesis and learned a lot about the topic of second language acquisition in general, as it is well-written and organized. The experiments are designed well, however, the sample size is small (especially given that the pre-test was done by a different group, so it is hard to assess what the individuals actually learned). However, I believe larger-

sized experiments are out of the scope of a Master's thesis, and the author tries to account for the issues as much as possible, so this limitation should not negatively affect the final assessment. Another possible improvement or future work could be designing similar experiments to measure the effectiveness of the developed software. The work is very interesting and the outcomes are of excellent quality, both in design and execution.

I recommend this work for the defense.

Questions:

- What was the motivation for measuring form recognition and meaning recall in the experiments and leaving out form recall and meaning recognition?
- In the experiments with *Whole* group - what do you suppose would be the effect of having longer episodes (longer show than Peppa Pig)? Conversely, what is the effect of segment length in the *Segments* group?
- What was the motivation for not running experiments without previewing in English, i.e. only in the target languages? Was it only because of the small number of participants?
- Do you think the ideas from your thesis could be used in language learning not only for humans, but for example large language models, e.g. that they can help to design efficient scheduling for the training examples?
- What do you suppose is the reason that both *Whole/Segments* groups do better than the *Control* group even on distractors, i.e. why do they have better prior vocabulary knowledge?
- On page 43, you say: "*Scores on distractor items were higher, ...*" Why? Because they are "easier" words? and the sentence continues: "*... as were form recognition scores compared to meaning recall, which was expected.*" Why? I would expect especially for Spanish-speaking participants in Italian to be able to infer the meaning of the word better, even without ever seeing the exact form of that word before.
- Do you plan similar experiments with languages that are more "exotic" for the target group? As you note, Italian was not well suited for some of the experiments for the participants with some prior knowledge of Spanish (or other Romance languages) due to a large number of cognates.

- Do you plan similar experiments using the software developed in the thesis, for example, how beneficial is using the spaced — in comparison to only watching the videos or a combination of both?

Prague, 27.8.2024

Signature: