

Thesis topics -

by Annemarie Verkerk

Please feel free to contact me at any point if you would be interested in writing your BA/BSc/MA/MSc thesis on a topic that I could advise on:

- linguistic typology
- historical linguistics, especially phylogenetics
- corpora-based cross-linguistic or contrastive studies

If you have no idea what such a project would look like, have a look at my main/research page or at the short list that follows:

Typology projects

Typology projects may involve existing typological databases, see here for an overview:

<https://linguistic-typology.org/databases/>

<http://humans-who-read-grammars.blogspot.com/p/free-online-linguistics-databases.html>

<http://languagegoldmine.com>

However, typology projects may also involve coding information about one or more specific variable(s) in a sample of languages, for which you will need access to grammars and other resources. Not to worry, because many people (most importantly Harald Hammarström) have digitized a lot of these materials over many years, so that would be available to you.

A typology project could also entail a detailed analysis of a phenomenon in an individual language, taking a typological perspective. For instance, asking the question how does your analysis of that phenomenon impact the analysis of other languages with similar characteristics?

A more specific lists follows:

- **Universals in typology.** If you need inspiration regarding the myriads of things that typologists have been claimed to be universal (i.e. hold across all languages in the world) in the past, consult the Konstanz Universals Archive: <https://typo.uni-konstanz.de/rara/category/universals-archive/>
There are over 2000 universals listed here that could be investigated further, either from a gradient-based or diachronic perspective, or from a family-based perspective (using phylogenetic comparative methods for instance, see below).
- **Linguistic typology using phylogenetic comparative methods.** I have done quite a bit of this myself (see main/research page). Using this set of methods, typological variables are modelled to change on the branches of a phylogenetic tree (language family tree). This allows us to answer all kinds of interesting questions, such as:
 - o Are two or more typological variables correlated or not?
 - o What was the ancestral state of the typological variable(s) for one or more Proto-languages?
 - o At what rate does diachronic change process? Comparing variables, is a variable faster or slower compared to another variable? (Ideas that have to do with diachronic stability.)
 - o Does change in one variable trigger change in another variable?

Since I have worked on most of these types of questions before, please see the main/research page for inspiration. Just one example for the latter type of question. Iker Salaberri has worked (in his BA thesis, actually) on word order change, comparing subordinate clauses to main clauses, discussing which clausal word order changes first: that of subordinate clauses or that of main clauses? Does change in one trigger change in the other? Evidence has been brought to bear on this question from a cross-linguistic perspective, but not from a cross-linguistic perspective that also incorporates language change across a wider family such as Indo-European. A family-based study of word order change in subordinate vs. main clauses would be an excellent thesis topic.

- **Negation.** I have worked on negative existential predicates with several collaborators (see main/research page). Strategies for negative existential predicates (i.e. saying things like '*There are no tame zebras*') evolve in interchange with standard negators (such as *not*, *nicht*) for verbal predicates. These standard negators in turn may change through the Jespersen's Cycle. Van Gelderen (to appear) has written about the interaction these and other cycles in morphosyntax. A study on their interaction in a single language family would be highly interesting. The Oceanic/Austronesian language family shows interesting variation in this respect.
- **Semantic typology.** This is the branch of typology that studies meaning, mostly on the basis of words and their denotation. This is an interesting field, because there are some semantic fields (see for a list the Concepticon, <https://concepticon.cldd.org>, which have been widely investigated from a cross-linguistic perspective: motion, kinship, color terms, and spatial relations, for example. Others remain unexplored. There is a lot of interesting work to do in this area, both from a survey-based perspective and from an experimental perspective. There are also databases like the Database of Cross-Linguistic Colexifications (<https://clics.cldd.org>) that can be used to work on this domain.
- **Grammatical gender.** Another topic on which I have personally worked is grammatical gender in Bantu, again with the specific focus on diachronic change. This work will be expanded in the future, but if this topic interests you, there would be options for data gathering and analysis in the same vein, or a focus on gender assignment in Bantu or another language family.

Phylogenetic inference (aka tree building)

This very short section is concerned with modern methods for inferring relationships between related languages. There are a lot of thesis topics to be found in this area, both from the perspective of formal phylogenetic modeling of cross-linguistic data, to comparing different types of data and methods (including character-based, distance-based, and network methods), and to answering more specific questions for instance regarding homelands, dating, punctuation of change, etc. Any thesis idea welcome, and I am happy to assist.

Corpora-based projects

As you may know from my classes, we have been building a large cross-linguistic parallel corpus in order to investigate information structure/management. I would welcome any thesis project that makes use of cross-linguistic corpora to study the process of translation (for that, you may also find supervisors in Prof. Teich's group). BUT I would be even happier with projects that use corpora NOT to study translation, but to study morphosyntax or even phonology from a cross-linguistic perspective (for the latter you may consider a corpus of audio books).

Corpora are inherently wonderful to study variation. And variation is often ignored/down-played in comparative work on morphology/syntax/morphosyntax. So many of the things one can think of here are new and will be innovative. Some examples:

- **Negation, again.** All Germanic languages have variation in their marking of negative existential predicates such that they may use the standard negator and the negative quantifier. English is a case in point:

<i>There are no tame zebras.</i>	<i>no</i> = negative quantifier
<i>There aren't any tame zebras.</i>	<i>aren't</i> = verb + contracted standard negator; + quantifier <i>any</i>

However, 1) they all do this a bit differently for negative existentials but most importantly, 2), the use of the negative quantifier in other predication domains (possessive, locative, equational, adjectival) is really varied across Germanic. A corpus-based approach even comparing the predicative usage of the negative quantifier in only a few languages would already be highly interesting.

- **Word order.** Levshina (2019), in her paper "Token-based typology and word order entropy: A study based on Universal Dependencies", has presented evidence for interactions between cross-linguistic and language-internal variation for many different languages. Many word orders have been neglected in typology, especially those that display significant language-internal variation. Examples are the order of object and oblique noun phrases, the order of the adverb and verb, and differences regarding clausal word order of full noun phrases vs. that of pronouns. That doesn't make these word orders any less interesting to investigate - so, good thesis topics here, too.
- **Morphological complexity.** Complexity is an intriguing and 'hot' topic, but it's really tough to implement concretely. For morphological complexity, there are several algorithms/variables that could be compared. The oldest one is morpheme-word ratio, famously introduced by Greenberg (1960). Another was implemented by Ehret (2017) in her article "An information-theoretic approach to language complexity: variation in naturalistic corpora". A comparison of a set of measures would be a cool thesis topic. Another comparison could be on the level of register in a smaller set of languages, i.e. to look if differences in morphological complexity exist across registers. Yet another possibility would be working on methods to extract morphology automatically in a cross-linguistic-informed sense, in line with work by Bernhard Wälchli and NLP.

And more

- In recent years, **cross-cultural economics and psychology** have started to explain economic and psychological behaviour in terms of characteristics of languages. Some famous examples are:

Keith Chen: Could your language affect your ability to save money?

<https://www.youtube.com/watch?v=lw3YTbubyjl>

Davis & Reynolds (2018) "Gendered language and the educational gender gap"

There has been follow-up work too, for instance Roberts et al. (2015), "Future tense and economic decisions: controlling for cultural evolution". Here is a blog post I wrote on this topic a few years ago:

<http://humans-who-read-grammars.blogspot.com/2017/03/spurious-correlations.html>

An excellent topic for a thesis would be a review of this work; recommendations on how this work should be interpreted, and replication.

- Of course, there are many more thesis topics possible. I just highlighted these to give you an idea of possibilities. I hope they inspire you!