Simulating language history

-

Inflectional change in the German verb system

Marisa Delz, Benjamin Layer, Sarah Schulz, Johannes Wahle[*]

Seminar für Sprachwissenschaft, Eberhard Karls Universität Tübingen, Wilhelmstrasse 19,
Tübingen, 72074, Germany
{marisa.delz; sarah.schulz; johannes.wahle}@student.uni-tuebingen.de
b-s-layer@gmx.de

October 7, 2011

Linguists working on the diachronic development of German observe a change in the inflectional verb system between Middle and Modern High German. In part due to the thinness of information it is difficult to determine the factors influencing this development. We investigated this with a model which simulates the change in the verbal morphology using as its basis just the linguistic observations from Middle and Modern High German. Neither intermediate stages nor other proposed interpretations of this change were taken into account. This is the first application of an Iterated Learning Model (ILM) such as Kirby's (Kirby & Hurford, 2002) to simulate change phenomena in real German language, so it also provides a test of the quality of the ILM as an explanatory model for language change. Our model could also be applied to other languages.

The phenomenon we address is the simplification of the German verb inflection system over time. In Middle High German the past tense form was formed either strong, weak, or irregular. The strong verbs are divided into seven ablaut classes. Each of these classes is based upon a change pattern which allows us to determine which verb belongs to which class. There were therefore in Middle High German two classes with regularities, the weak and the strong class, and one without regularities, the irregular class (Hennings, 2003; Paul, 1989). In our analysis in Modern High German the former strong and irregular classes have merged, leaving only two classes: the regular – formerly the weak – and the irregular verbs. The regular verbs form the major category, whilst the irregular verbs are far fewer. We claim that the loss of the ablaut classes was not a spontaneous occurrence in the language but a regular evolutionary process. There is a statistic which illustrates this reduction. In Old High German, there are 349 irregular verbs, in Middle High German 339, and in Modern High German just 169 irregulars left (c.f. Nübling, 2008, p. 213). We assume that the change in the German verb system is dependent upon the frequency of usage and the process of overgeneralization. Irregular verbs have a higher frequency of usage than regular verbs (c.f. Nübling, 2008, p. 57): most of the irregular verbs belong to the basic vocabulary of the language (c.f. Eisenberg, 2009, p. 440). Although they need to be saved in the mental lexicon, it is worthwhile for a learner to memorize the morphological forms of these verbs. The regular verbs are less frequent; it is therefore more economic to form them via a rule (c.f. Rohr, 1999; Pinker & Prince, 1994).

---

[*]Author names appear in alphabetical order.

Our account of this phenomenon is based on the idea of iterated learning. The central element of every ILM is a learning algorithm allowing generalization over learned forms. There are two competing mechanisms: the holistic one, just looking at the data and saving it, and the generalizing one, comparing the data and calculating a probability for each inflectional form. To model this learning, different algorithms have been used. The most frequent ones are based on neuronal networks (c.f. Kirby & Hurford, 2002; Smith, Kirby, & Brighton, 2003) and grammar induction (c.f. Kirby & Hurford, 2002). The population in our model at any one time consists of just one *speaker* teaching the language and one *learner* acquiring it. We start with a realistic distribution of verb inflection classes and update it every generation. Every subsequent speaker will teach the class assignments he has learned including any inflection class that was learned by overgeneralization. The crucial term in this context is that of the *bottleneck*. The bottleneck describes the imperfect transmission between the language of the speaker and the one of the learner, i.e. the language as it is used by a population and the language internalized by a new speaker (the next generation). Each language state adopts two roles: in the learner it is the acquired language; in the following generation it becomes the teacher language. Changes are thus passed on cumulatively.

Since the change in German took place over many generations, the ILM approach is an appropriate framework for the implementation. The model includes three main parameters: the frequency of the verbs, the degree of restriction at the bottleneck and the number of iterations, i.e. the number of generations. The overgeneralization mechanism relies on the frequency of the inflection classes, so the frequency of a verb has a strong influence on its evolution. We therefore keep the verb list and the frequency values constant over the generations. We start off with a database containing 327 verbs constructed on the basis of the digitized Lexer (Burch, 2002) and Ruoff (1981). On the basis of Burch (2002) we strove to construct a maximally realistic distribution of the ablaut classes. The assignment to an inflection class is the only dynamic element in this model. Verbs were selected at random out of the pool of 327 for transmission from the teacher to the learner. A bottleneck size is defined by the number of verbs taught. A value of 100% means that exactly 327 randomly chosen verb tokens were taught in a given generation.

(a) Example output with a bottleneck size (= BNS) of 50%   (b) Averaged over 100 outputs with a bottleneck size of 50%
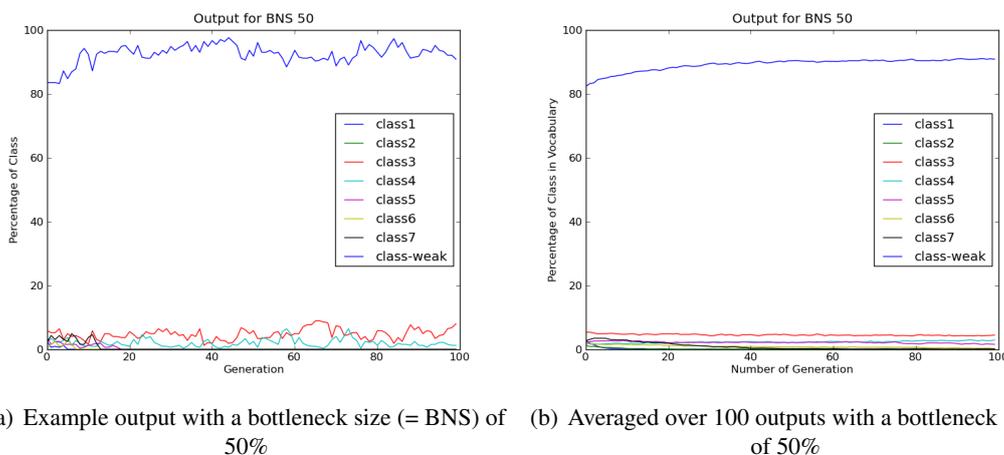
Figure 1: *output graphs showing distributions of class assignments over generations*

Due to the the size of the database the seven ablaut classes are all fairly small, so that no clear boundaries

can be seen in our output at the beginning. Nonetheless, the graphs show the general development of the German verb system, the increase in the weak verb class, and the decrease of the other classes. Figure 1(a) shows a typical, randomly chosen output of a single run of 100 generations. This result would predict that Modern High German should contain no more verbs belonging to classes 1, 2, 5, 6 and 7. In fact modern German contains such verbs. This may be because our model uses only one learner per generation and other language conserving factors (e.g. written documents, school grammars, etc.) were not taken into account. Each individual result is however subject to probabilistic fluctuation. To eliminate these we averaged over 100 different outputs, see Figure 1(b).

The size of the bottleneck also influences the resulting graph. Increasing the bottleneck size results in more stable verb classes; reducing it leads to a rapid disappearance of classes. Bottleneck sizes between 50% and 100% created realistic outcomes. The frequent verbs persist in the strong inflection classes while the rare ones tend to change their classes.

The ILM seems to offer an adequate framework to replicate the development in the structure of the German inflectional system over the last 800 years. The learning and generalization mechanism we use is based on the simple assumption that there is a tendency to inflect unknown verbs in the same way as most of the others. The pressure of learning through a bottleneck triggers changes allowing forms to be attracted from smaller classes to the bigger ones. The complex inflection system based on verb classes disappears and changes into a system with only weak and strong forms, where the strong forms are remnants of the Middle High German strong inflectional class. Our model successfully simulates the change in the German language from Middle High German to Modern High German and supports the claim that the change is due to the frequency of usage and the process of overgeneralization. The fact that we used a nearly realistic database for our model supports the claim, too. The model we constructed can be used as a basis for further research on this topic.

References

Burch, T. H. (Ed.). (2002). *Mittelhochdeutsche Wörterbücher im Verbund.* Stuttgart: Hirzel [u.a.].

Eisenberg, K. R., Peter ; Kunkel-Razum (Ed.). (2009). *Der Duden in zwölf Bänden* (Vol. 4: Duden - Die Grammatik : unentbehrlich für richtiges Deutsch). Mannheim: Dudenverl.

Hennings, T. (2003). *Einführung in das Mittelhochdeutsche.* Walter de Gruyter.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer Verlag.

Nübling, D. (2008). *Historische Sprachwissenschaft des Deutschen: eine Einführung in die Prinzipien des Sprachwandels* (2. überarb. Aufl. ed.). Narr.

Paul, H. (1989). *Mittelhochdeutsche Grammatik.* Niemeyer.

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima (Ed.), *The reality of linguistic rules.* Amsterdam [u.a.]: Benjamins.

Rohr, W. G. (1999). *Einführung in die historische Grammatik des Deutschen.* Hamburg: Buske.

Ruoff, A. (1981). *Häufigkeitswörterbuch gesprochener Sprache: gesondert nach Wortarten, alphabetisch, rückläufig alphabetisch und nach Häufigkeit geordnet.* Max Niemeyer Verlag.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial Life*, *9*, 371-386.