The Development of Adjective Morphosyntax in High German: Using Information Theory to Quantify Claims about Language Change*

David M. Howcroft, Cynthia A. Johnson, Rory Turnbull,
Rachel Steindel Burdin, & Tsz-Him Tsui
The Ohio State University, Department of Linguistics

caj@ling.ohio-state.edu
April 26, 2013

1 Introduction

1. Middle High German (MHG) adjectives have strong and weak forms (see Tables 3 and 4 in Appendix A for paradigms).

2. New High German (NHG) adjectives have strong and weak, and also "mixed" forms (see Tables 5, 6, and 7 in Appendix A).

3. Synchronically, grammar handbooks describe alternation between adjective forms as a product of relevant morphosyntactic information being available somewhere in the noun phrase as a whole.

   (a) NHG: Durrell (2002:118) claims "the underlying principle which governs the use of the strong and weak declension is that the fuller 'strong' endings are used when there is no determiner preceding the adjective with an ending which indicates the case, gender or number of the noun as clearly as possible."

   (b) MHG: Walshe (1974:17) finds that MHG strong and weak forms are distributed "according to much the same rules as New High German but with less fixity of usage."

4. The availability of information is also (somewhat) apparent in terms "strong" and "weak": like strong nouns and verbs, strong adjectives give more morphological information.

*We are grateful to Brian Joseph, Andrea Sims, Anna A. Grotans, and Tonya Kim Dewey, along with the OSU Changelings Discussion Group and the audience at the 2013 MLK Jr. Symposium at OSU for helpful discussion and comments. Any remaining errors are our own.

5. Explanations of morphological change (e.g. shift from the bipartite definiteness system in MHG to the tripartite system in NHG) often couched in terms like "clarity" and "redundancy":

   - "[The development of a mixed system] is motivated by the fact that the alternation is serving a morphological "purpose"—of more clearly marking plural forms." (Hock 1991:446)
   - "[Mańczak's second tendency and Kuryłowicz's first "law"] reflect tensions present in languages in general: respectively the need to have redundancy for clarity and the desire to eliminate unnecessary or unmotivated redundancy." (Joseph 1998:365)
   - "Adaptive innovations: The deductive innovations which arise in response to a perceived inadequacy of a grammar to carry the semantic content the speakers wish it to express can be categorized according to the communicative function they intend to serve." (Andersen 1980:8–9)

6. These claims about the synchronic status of the High German adjective forms, and diachronic variation are intuitionistic: while grounded in theories of linguistic typology and basic assumptions about cognition, they lack an empirical basis.

7. In this paper, we provide an empirical assessment of these claims, and find evidence consistent with the synchronic claims; however, the diachronic results present a more complicated picture.

2 Background

1. In MHG, weak forms were used with any word that declines like the definite article, even if not definite in meaning: *manch-*, 'some'.

2. Strong forms were used elsewhere, even if article is definite in meaning: *mîn*, 'my' (MHG).[1]

3. Change in usage in NHG:

   (a) Strong form used when no article or an article that doesn't decline

   (b) Weak forms still used with definite article

   (c) New "mixed" forms (literally a mix of the strong and weak forms) used with the indefinite article.

4. Grammarians find that case, number, and gender information is usually available *somewhere* in the NP, if not fully specified by the form of the determiner or adjective alone, cf. Durrell's prescription above.

[1] Note that these are the usage rules prescribed by grammar books; actual usage "on the ground" was quite variable, particularly for MHG.

5. Formal syntactic accounts of the adjective distribution in Germanic model the information-marking as redundancy-free:

   - "…we might expect only one element of the NP to exhibit it. The natural choice, of course, is the first element, so that the hearer can immediately identify the function without having to wait until the end of the NP (although any other element of the NP or even the sentence could, of course, carry this same information)" (Esau 1973:139)
   - "The generalization drawn from this is that the adjective and the article compete for the same marker. If the marker is not present on the article, it must appear on the adjective" (Biskup 2007:241)
   - "…in the ideal case, this information [case, number, gender, and definiteness] is conveyed unambiguously and nonredundantly [within the NP]" (Zwicky 1986:975), criticizing previous accounts)

6. The goal of the present study is to determine whether the strong forms demonstrably carry more information than the weak forms, compensating for the differences in the forms of the articles.

7. The tools of information theory can be used to address this question.

3 Information Theory

1. Information theory (Shannon 1948) provides mathematical measures which can be used to compare the amount of information conveyed by morphological forms.

2. We can use these measures to compare across NPs synchronically and across adjective inflectional systems diachronically.

   (a) Synchronically, we can compare the information content of the strong, weak, and mixed adjective forms with each other.

   (b) Diachronically, we can compare the MHG and NHG systems with each other.

3. Information theory was developed in telecommunications research as a precise means of quantifying the information conveyed by a message (e.g. a word) given its context (e.g. the preceding words in a sentence). In particular, the information conveyed by a signal is inversely related to its probability within the system (e.g. the English language).

4. When an item is entirely predictable it conveys no information. This is why we can say *a bird in the hand* as a proxy for the full expression *a bird in the hand is worth two in the bush* while conveying the full metaphoric meaning of the expression.

5. In contrast, highly unpredictable items convey more information.

6. *Entropy* is a measure of the overall unpredictability of a system, and thus represents the distribution of information within a system as a whole.

7. A system in which every outcome is equally unpredictable (e.g. rolling a fair die) exhibits maximum entropy.

8. By contrast, a system with a very skewed set of outcome probabilities (e.g. a loaded die) has lower entropy overall, i.e. the outcomes are less surprising in general because some of the outcomes are more probable than others.

3.1 Information Theory and Linguistics

1. Information theory has been applied to linguistic phenomena to characterize the organization of the grammatical system and the variation observed.

2. Recent information theoretic approaches to other linguistic phenomena:

   - Phonological structure (Goldsmith & Riggle 2012)
   - /t/- and /d/-deletion in English (Jurafsky et al. 2001)
   - Phonological contrast and neutralization in sound change (Hall 2009)
   - The merger of tones in some varieties of Cantonese (Tsui 2012)
   - The presence or absence of *that* in reduced relative clauses (Levy & Jaeger 2007)
   - The organization of paradigms and morphological complexity (Sims 2011, Ackerman et al. 2009, Malouf & Ackerman 2010)

3. With regard to the Germanic adjective system:

   (a) Synchronically, we can compare the entropy of the strong, weak, and mixed adjective forms with each other.

   (b) Diachronically, we can compare the entropy of the MHG system with the entropy of the NHG system.

4 Methodology

4.1 Calculating Entropy

1. We begin by examining MHG and NHG synchronically.

2. Measuring the conditional entropy of an adjective form given the article form allows us to evaluate claims that, e.g. the strong endings are compensating for the absence of clear morphosyntactic expression in the preceding context.[2]

[2]We make the simplifying assumption that the preceding context is one of EIN, DER, or an indeclinable article, denoted as ∅.

3. First, we calculate the conditional probability of each ending given the article.

4. For example, the strong form is licensed by the MHG indeclinable *solch*, but the form of the adjective cannot be predicted given *solch*.

5. There are twelve paradigm cells, and *solch* takes the same form in each of them.

6. As an example, the probability of the strong ending *-iu* given *solch* is $P(\text{-}iu|solch) = \frac{1}{12}$, since *-iu* only occurs in one paradigm cell.

7. The definition of the entropy H of a system X is:

$$H(X) = -\sum P(x_i)\log_2 P(x_i).$$

8. For the portion of the adjectival system consisting of adjectives following *-er* (e.g. following *einer*, *der*), we have:

$$\begin{aligned} H(X) &= -(P(\text{-}e|\text{-}er)\log_2 P(\text{-}e|\text{-}er) + P(\text{-}er|\text{-}er)\log_2 P(\text{-}er|\text{-}er) + P(\text{-}en|\text{-}er)\log_2 P(\text{-}en|\text{-}er)) \\ &= -(\frac{1}{6}\cdot\log_2\frac{1}{6} + \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{3}\log_2\frac{1}{3}) \\ &= -(-0.431 - 0.500 - 0.528) \\ &= 1.459 \end{aligned}$$

9. We averaged across the entropy of each possible article ending to arrive at an entropy value for the system as a whole.

10. Thus we determine the predictability of adjective forms given preceding context (i.e. indeclinable form, EIN, or DER) and compare average entropy of strong and weak forms in MHG and strong, weak, and mixed forms in NHG to evaluate the synchronic claims of the grammarians.

11. In order to evaluate the claims about the relative contribution of information from the article and the adjective, we also calculate the predicability of the articles given each adjective ending.

12. For an entropy value $H(A|B)$, a higher value means that B carries less information about A.

13. This approach also allows us to compare the average entropy of the bipartite (strong and weak) adjectival system of MHG with that of the tripartite (strong, weak, and mixed) system in NHG.

14. We use dictionaries to determine the relative probabilities of noun gender (see Table B in Appendix B).

    (a) MHG: the BMZ (Benecke, Müller, & Zarncke dictionary)

(b) NHG: West's `germanstudies.org.uk` dictionary

15. However, dictionaries cannot tell us relative token frequencies of articles and of case forms; for the purposes of this study, we assume that they are equiprobable.

16. We only address the singular paradigms. Neutralization of the plural in NHG makes the comparison of plural paradigms in NHG with MHG non-trivial.

5 Results & Discussion

5.1 Synchronic Comparisons

	Strong	Weak
$H(\text{ADJ}\vert\text{ART})$	1.34	0.19
$H(\text{ART}\vert\text{ADJ})$	1	2.06

Table 1: Average entropies for MHG

Key		
Higher entropy	$\rightarrow$	Less informative
Lower entropy	$\rightarrow$	More informative

1. The results for MHG are shown in Table 1.

2. The grammarians are correct in characterizing the strong endings as more informative than the weak endings (1.34 vs. 0.19 in the first row).

3. They are also correct in that the more informative strong endings occur with the less informative articles (1 vs. 2.06 in the second row).

	Strong	Mixed	Weak
$H(\text{ADJ}\vert\text{ART})$	2.24	0.23	0.24
$H(\text{ART}\vert\text{ADJ})$	0	1.16	1.79

Table 2: Average entropies for NHG

4. Table 2 show the results for NHG.

5. The pattern observed in MHG holds for strong and weak forms in NHG.

6. The mixed forms are approximately as predictable as the weak forms (0.23 vs. 0.24).

7. However, the mixed form is more informative about the article (1.16 vs. 1.79) than the weak form, suggesting that the mixed forms do indeed bear more information than the weak forms.

8. Note that the article is completely predictable from the strong form in NHG (0 vs. 1.16).

5.2 Diachronic Results

1. The average entropy of the MHG system is 1.32.
2. The average entropy of the NHG system is 0.93.
3. Entropy has decreased.
4. A view of language change as grammar simplification might predict that entropy should always decrease in later stages of the language.
5. However, even if entropy does decrease, as it has in this case, this does not necessarily mean that all morphological change is simplification.
6. Compare the loss of case systems in English (a simplification); however, word order became much more fixed (a complication: word order is now stipulated).
7. Rather, this decrease in entropy might instead be a reaction to changes in other parts of the NP or High German grammar as a whole (where entropy may have increased).

6 Conclusion and Future Directions

1. The claims of the grammarians regarding the adjectival systems are quantifiably correct: strong forms are more informative than the weak and mixed forms.
2. The strong forms pattern with the less informative articles; the mixed and weak forms, with the more informative articles.
3. We see a decrease in entropy in the adjectival system from MHG to NHG; this might have been in reaction to increase in entropy elsewhere.
4. A corpus study will help answer this question and also allow us to control for relative frequencies of case and article contexts.
5. Another direction for future work is determining why strong and weak forms are distributed the way they are in the NHG mixed paradigm.[3]
6. Mixed forms in NHG might thus represent some optimal solution, given other pressures on the system; we can test this using information theory, as we have demonstrated with the synchronic status.

[3]Wright (1917:77) describes variation in use of forms with *ein* in MHG: "Ein and the possessive pronouns are followed by the strong form in the Nom. and Acc. singular; by the strong **or** weak form in pl. and the Gen. and Dative singular". This variation suggests a change in progress leading to the modern mixed paradigm.

A MHG and NHG Adjective Paradigms

MHG Declensions

Table 3: Middle High German: Strong (indefinite, no article) adjective declension- singular

Case	Masc “a/such a good day”	Neut “a/such a good word”	Fem “a/such a good gift”
Nom	einer/solch guoter tac	einez/solch guotez wort	ein/solch guotiu gâbe
Acc	einen/solch guoten tac	einez/solch guotez wortes	eine/solch guote gâbe
Gen	eines/solch guotes tages	eines/solch guotes wortes	einer/solch guoter gâbe
Dat	einem(e)/solch guotem(e) tage	einem(e)/solch guotem(e) wortes	einer/solch guoter gâbe

Table 4: Middle High German: Weak (definite) adjective declension

Case	Masc “the good day”	Neut “the good word”	Fem “the good gift ”
Nom	der guote tac	daz guote wort	diu guote gâbe
Acc	den guoten tac	daz guote wort	die guoten gâbe
Gen	des guoten tages	des guoten wortes	der guoten gâbe
Dat	dem(e) guoten tag	dem(e) guoten worte	der guoten gâbe

NHG Declensions

Table 5: Standard NHG: Strong (no article) adjective declension

Case	Masc “some good day”	Neut “some good word”	Fem “some good gift ”
Nom	etwas guter Tag	etwas gutes Wort	etwas gute Gabe
Acc	etwas guten Tag	etwas gutes Wort	etwas gute Gabe
Gen	etwas guten Tages	etwas guten Wortes	etwas guter Gabe
Dat	etwas gutem Tag	etwas gutem Worte	etwas guter Gabe

Table 6: Standard NHG: Mixed (indefinite) adjective declension

Case	Masc “a good day”	Neut “a good word”	Fem “a good gift ”
Nom	ein guter Tag	ein gutes Wort	eine gute Gabe
Acc	einen guten Tag	ein gutes Wort	eine gute Gabe
Gen	eines guten Tages	eines guten Wortes	einer guten Gabe
Dat	einem guten Tag	einem guten Worte	einer guten Gabe

Table 7: Standard NHG: Weak (definite) adjective declension

Case	Masc “The good day”	Neut “The good word”	Fem “The good gift ”
Nom	der gute Tag	das gute Wort	die gute Gabe
Acc	den guten Tag	das gute Wort	die gute Gabe
Gen	des guten Tages	des guten Wortes	der guten Gabe
Dat	dem guten Tag	dem guten Worte	der guten Gabe

B Gender counts

Table 8: Counts of nouns in each gender for MHG and NHG

Language	Masc	Neut	Fem
MHG	3343	1858	3227
NHG	5281	4429	4738

References

Ackerman, Farrell, James P. Blevins, and Robert Malouf. 2009. "Parts and Whole: Implicative Patterns in Inflectional Paradigms." In J.P. Blevins and J. Blevins (eds.), *Analogy in Grammar*. Oxford: Oxford University Press.

Andersen, Henning. 1980. "Morphological Change: Towards a Typology." In Jacek Fisiak (ed.) *Historical Morphology*. The Hague: Mouton de Gruyter.

Biskup, Petr. 2007. "Spellout and the weak, strong, and mixed declension in German and definiteness in Bulgarian." *Linguistische Arbeits Berichte*, 85, 241–261

Durrell, Martin. 2002. *Hammers German Grammar and Usage*, Fourth edition. London: Hodder Education.

Esau, Helmut. "Form and Function of German Adjective Endings." *Folia Linguistica*, 6:1, 136–145

Goldsmith, John, and Jason Riggle. 2012. "Information theoretic approaches to phonological structure: the case of Finnish vowel harmony." *Natural Language and Linguistic Theory*, 30, 859–896.

Hall, Kathleen Currie. 2009. *A Probabilistic Model of Phonological Relationships from Contrast to Allophony.* PhD Dissertation, The Ohio State University.

Hock, Hans Henrich. 1991. *Principles of Historical Linguistics*, Berlin: Mouton de Gruyter.

Joseph, Brian. 1998. "Historical Morphology." In A. Zwicky & A. Spencer (eds.), *The Handbook of Morphology*, Blackwell Publishers.

Jurafsky, Daniel, Alan Bell, Michelle L. Gregory & William D. Raymond. 2001. "Probabilistic relations between words: Evidence from reduction in lexical production." Joan L. Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam: Benjamins, 229–254.

Levy, Roger, and T Florian Jaeger. 2007. "Speakers Optimize Information Density Through Syntactic Reduction." *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*.

Malouf, Robert, and Farrell Ackerman. 2010. "Paradigms: The low entropy conjecture." Paper presented at Workshop on Morphology and Formal Grammar, Paris.

Sims, Andrea D. 2011. "Information Theory and Paradigmatic Morphology." Presented at *Workshop on Information Theoretic Approaches to Linguistics*, Boulder, CO: 15–16 July 2011.

Shannon, C. E. 1948. "A mathematical theory of communication." *Bell System Technical Journal*, 27, 379–423, 623–656.

Tsui, Tsz-Him. 2012. "Tonal variations in Hong Kong Cantonese: Interactions of acoustic distance and functional load." Presented at CLS48, Chicago IL.

Walshe, M. O'C. 1974. *A Middle High German Reader*. Oxford: Oxford University Press

Wright, Joseph. 1917. *A Middle High German Primer*, Third Edition. Oxford: Oxford at Clarendon Press.

Zwicky, Arnold M. 1986. "German Adjective Agreement." *Linguistics*, 24. 956–990.