Linguistic Assessment Criteria for Explaining Language Change: A Case Study on Syncretism in German Definite Articles

Remi van Trijp

Sony Computer Science Laboratory Paris

Paris, FRANCE

Abstract

The German definite article paradigm, which is notorious for its case syncretism, is widely considered to be the accidental by-product of diachronic changes. This paper argues instead that the evolution of the paradigm has been motivated by the needs and constraints of language usage. This hypothesis is supported by experiments that compare the current paradigm to its Old High German ancestor (OHG; 900–1100 AD) in terms of linguistic assessment criteria such as cue reliability, processing efficiency and ease of articulation. Such a comparison has been made possible by “bringing back alive” the OHG system through a computational reconstruction in the form of a processing model. The experiments demonstrate that syncretism has made the New High German system more efficient for processing, pronunciation and perception than its historical predecessor, without harming the language’s strength at disambiguating utterances.

Keywords

case syncretism; computational modeling; language change; German definite articles; Fluid Construction Grammar

1. Introduction

In his 1880 essay, Mark Twain famously complained that The awful German language is the most “slipshod and systemless” language of all, and “so slippery and elusive to grasp.” A brief glance at the literature on the German case system suggests that many linguists agree with the American author, even though

“The author is fascinated by the origins and evolution of language. In his research, he combines techniques from computational linguistics and A.I. to understand how grammatical systems may emerge and evolve. Van Trijp holds a PhD in computational linguistics from the University of Antwerp and has been an associate researcher at the Sony Computer Science Laboratory Paris since 2008.”
of course they speak of “idiosyncrasies” or “historical accidents” rather than sharing Twain’s harsh assessment. But what if the German case system were not the accidental by-product of diachronic changes as is often assumed? Are there linguistic forces at play that are not yet fully appreciated in the field, but which may explain the German case paradigm—and therefore potentially many other linguistic phenomena as well?

In this paper I argue that there are indeed such forces, and I present a novel research method for revealing them. This method is applied to one of the biggest mysteries of the German language, namely its system of definite articles, whose forms are marked for case, number and gender. The paradigm is notorious for its syncretism, that is, the same article can be mapped onto multiple different functions. For instance, the article der ‘the’ can be used as a determiner (or pronoun) for nouns that are (a) nominative-singular-masculine, (b) dative-singular-feminine, (c) genitive-singular-feminine, or (d) genitive-plural. The entire paradigm is shown in the right of Fig. 1.

The fact that the paradigm has syncretic forms is not surprising in itself, of course, as syncretism is a widely attested property of inflectional languages (Baerman, 2009). The real puzzle is why these syncretisms came about, especially when taking a historical perspective. To the left of the New High German system of definite articles, Fig. 1 also shows the paradigm in two earlier stages of its evolution: Old High German (OHG; 900–1100; Wright, 1906) and Middle High
German (MHG; 1100–1500; Wright, 1916). What immediately jumps to the eye in the OHG paradigm is that it had a much more transparent mapping between form and function. For instance, where New High German (NHG) has one syncretic form *die* for nominative and accusative plurals, OHG had a three-way gender distinction between masculine, neuter and feminine. In total, the OHG paradigm counted twice as many distinct forms as the current system. So why did the speakers of German allow this more transparent system to crumble down to its current form?

1.1. Syncretism by Accident

The more general question of why languages may lose their case morphology has been addressed in numerous historical studies (e.g., Barðdal, 2009, Barðdal and Kulikov, 2009, Blake, 2001, Boas, 2009, Harbert, 2007). Unfortunately, despite the fact that “the loss of morphological case in the Germanic languages has been subject to substantial research for a long time in linguistics, [...] there is no general consensus on its causes” (Barðdal, 2009: 123). A general tendency is to consider the loss of case distinctions and the subsequent increase in syncretism as the result of *erosion* (Barðdal and Kulikov, 2009, Blake, 2001, Kulikov, 2009), which “reaches its endpoint when the case marker is lost” (Heine, 2009: 459).

As suggested by the term “erosion,” which evokes the process of soil, rock or land being worn away by natural forces such as water and wind, there is an implicit assumption that language change is similarly governed by natural laws that mechanically apply if certain conditions are met. One example is *phonological erosion*, as can be observed in the OHG dative-singular-feminine form *dëru*, which has shortened into *der*—and thus collapsed with the form of the nominative-singular-masculine definite article. Blake (2001: 169) argues that such “reductive phonological changes can all but destroy a morphological case system.”

Most historical linguists nowadays agree that syncretism is almost never the outcome of a single pressure, but rather the “result of a complex interplay of several mechanisms” (Barðdal and Kulikov, 2009: 474). Indeed, many other eroding forces have been proposed in the literature, ranging from language contact to the typological shift from synthetic to analytic languages (see Barðdal, 2009, for a critical review). Whatever the combination of pressures proposed for explaining specific instances of language change, case syncretism is often assumed to be the accidental by-product of these pressures, and German case syncretism is typically analyzed along these lines (Baerman, 2009: 229).
1.2. An Alternative Hypothesis

In this paper, I explore an alternative to the syncretism-by-accident hypothesis, namely that the evolution of the German definite articles is driven by the needs and constraints of language users. This hypothesis subscribes to an increasingly popular line of research that can be grouped together under the umbrella term of “usage-based approaches to language” (Bybee and Beckner, 2010). What sets this paper apart from previous studies is that it tries to reveal the driving forces of language change by comparing two historical stages of a language in terms of linguistic assessment criteria, of which we can plausibly assume that they form an integral part of the linguistic experience of each individual language user.

In this paper, I make such a comparison between the current system of definite articles in German (henceforth called the “NHG system”) and its Old High German predecessor (the “OHG system”). The striking result is that the increase in case syncretism has made the NHG system more efficient than the OHG system for processing, articulation and perception—without, however, harming the language’s strength at disambiguating utterances. Moreover, the assessment criteria also explain the choice of which specific articles collapsed into syncretic forms: the OHG forms that are found to be “less efficient” are also the ones that have actually undergone change in the history of German.

2. Experiments

How can we compare the linguistic experience of present-day German language users to that of their forefathers? With no speakers of Old High German left, it is impossible to conduct neuro- or psycholinguistic experiments to answer this question. However, the development of new tools in computational linguistics and AI have made it possible to come up with a solid basis for comparison.

2.1. A Processing Model for Old High German

Historical linguists have been able to reconstruct the linguistic inventory of Old High German through careful analysis of texts (Wright, 1906). These inventories can be exploited for also reconstructing a (computational) processing model for Old High German that is capable of analyzing utterances into a meaning (i.e. parsing, as performed by a listener) and of verbalizing a meaning into an utterance (i.e. producing, as performed by a speaker).

Most contemporary studies in cognitive modeling implement computational processing models using probabilistic methods that can be trained on corpus data (e.g., Levy, 2008). Unfortunately, such models cannot be used for Old High Ger-
man because the few OHG texts that have survived (Schützeichel, 2006: 9–19) are not annotated and too genre-specific to be an authentic reflection of actual language usage. The most robust solution to sparse data (or even no data at all for reconstructed languages such as Proto-Indo-European) is to implement models of **deep language processing**. The experiments reported in this paper feature deep language processing models in Fluid Construction Grammar (FCG; www.fcg-net.org; Steels, 2011, Steels, 2012).

This paper does not require any background knowledge in the computational details of the reconstruction; readers who are nevertheless interested in how everything “works” are kindly invited to check the linguistic facts behind the implementation in the Appendix and an interactive web demonstration of parsing and production at www.fcg-net.org/demos/german-case.

### 2.2. Experimental Set-up

The experiments operationalize a speech community as a *population of autonomous artificial agents* (henceforth called *agents*) that engage with each other in locally situated communicative interactions (called *language games*; Steels, 1995). Each agent is a model of an individual language user, and its linguistic behavior is influenced by its social and communicative goals, knowledge and beliefs, past experiences, appreciation of the speech setting, and cognitive and sensorimotor constraints. In other words, the methodology allows us to operationalize, inspect and control all aspects of a usage-based language model. More specifically, two different populations of agents are initialized with a population size of $N = 10$ for each. The members of the first population (henceforth called *OHG agents*) are endowed with the processing model of Old High German, while the members of the second population (*NHG agents*) are provided with the processing model of New High German.

### 2.3. Communicative Task

In each language game, two agents are randomly drawn from the same population. One agent is assigned the role of the speaker, and the other the role of listener. The speaker has to describe a certain state-of-affairs in the current speech setting using a declarative utterance. The listener then tries to comprehend the utterance and interpret its meaning. The agents are assumed to reach communicative success in each interaction. That is, even if the description of the speaker is ambiguous, the speech setting is assumed to be sufficiently clear to allow the listener to infer the intended meaning from the context.

Due to the lack of reliable frequency data for Old High German, the experiments focus on utterance *types* instead of *tokens*. The agents are capable of
processing three basic utterance types that are commonly found in both OHG and NHG texts:

1. Ditransitive: NOM—verb—DAT—ACC
   e.g., *Die Kinder gaben der Frau die Zeichnung.*
   ‘The children gave the woman the drawing.’
2. Transitive (a): NOM—verb—ACC
   e.g., *Die Frau sah den Mann.*
   ‘The woman saw the man.’
3. Transitive (b): NOM—verb—DAT
   e.g., *Der Mann hilft der Frau.*
   ‘The man helps the woman.’

The meanings of the utterances consist of a verb (e.g., ‘to help’), its participant roles (e.g., a ‘helper-role’ and a ‘beneficiary-role’), and its arguments (e.g., a ‘man’ and a ‘woman’). Meanings are represented using a first-order predicate logic:

(1) \[\exists ev, x, y : \text{help}(ev), \text{helper}(ev, x), \text{beneficiary}(ev, y), \text{man}(x), \text{woman}(y)\]

The meanings of the arguments always correspond to distinct lexical forms for singular and plural (e.g. *Mann* ‘man’ vs. *Männer* ‘men’), which, however, are either unmarked for case or whose case marking is ignored (see the Appendix). All meanings are provided by a meaning generator, which ensures that the combination of arguments in the meaning that needs to be expressed by the speaker is always unique among the dimensions of number and gender, which yields 216 unique utterance types:

(2) NOM.SG.M Verb DAT.SG.M ACC.SG.M
    NOM.SG.M Verb DAT.SG.F ACC.SG.M
    NOM.SG.M Verb DAT.SG.N ACC.SG.M
    etc.

In transitive utterances, there is an additional distinction based on animacy for nouns in the object position of the utterance, which yields 72 types in the NOM-ACC configuration and 72 types in the NOM-DAT configuration. Together, there are 360 unique utterance subtypes. The genitive case is not considered in the experiments because it is not part of the most productive argument structure patterns in German (Shrier, 1965).
3. Linguistic Assessment Criteria

The following subsections introduce the linguistic assessment criteria that are tracked while the agents play language games with each other. The results of all the individual experiences are then aggregated and extrapolated on a scale between zero and one, for which holds that the higher the score, the better the performance. The criteria themselves are assumed to be cognitively plausible, but the research does not commit to a specific implementation.

In the remainder of this paper, I will use the term language system to refer to a particular grammatical paradigm of a language, such as its system of definite articles, tense-aspect system, and so on. I will use the term language to refer to the whole linguistic inventory (i.e. containing the lexicon and all language systems). Language systems may recruit various morphosyntactic means for expressing their paradigmatic distinctions. For instance, the English tense system employs the marker -ed on regular verbs for signaling that an event took place in the past (as in walk vs. walk-ed). I will refer to the morphosyntactic realization of a language system as a cue to the listener to help in the interpretation of utterances.

3.1. Cue Reliability and Disambiguation Power

The primary function of case (and hence of the case-marked definite articles in German) is to help the listener to identify “who did what to whom.” Ambiguous utterances that lead to multiple possible interpretations are assumed to require more interpretation effort on the part of the listener, who needs to perform additional inferences from the context to retrieve the intended meaning. The measure disambiguation power assesses how adequate a language is at eliminating different possible interpretations of an utterance, hence facilitating the listener’s interpretation task. The measure cue reliability assesses the disambiguation power when only the morphosyntactic forms of a single language system are taken into account. Intuitively speaking, we may expect language systems that are more transparent in how they map form onto meaning to have a higher cue reliability than systems that are full of syncretic forms, and languages with more reliable grammatical systems to display more disambiguation power than languages that have less reliable systems.

Measure

Both cue reliability and disambiguation power can be calculated based on utterance disambiguation, that is, based on the amount of utterances that can be disambiguated during parsing. This paper operationalizes utterance ambiguity as follows. Suppose that the listener has to parse the NHG utterance *Der Hund*
beißt den Mann ‘the dog bites the man.’ When parsing, lexical constructions add a meaning predicate for each noun and verb, as already illustrated in example (1) above, but this time the meaning contains variables (indicated with a question mark):

\[ (3) \]

When interpreting this meaning, the listener has to bind these variables to referents in the world. For instance, the variable \( ?ev \) needs to be bound to the specific instance of a bite-event that the speaker is talking about, the variable \( ?x \) needs to be bound to the specific participant who did the biting, and so on. Note, however, that the meaning of the event’s participant roles contains variables (\( ?x \) and \( ?y \)) that are different from the meanings of the participants (\( ?a \) and \( ?b \)), which captures the fact that lexical constructions only introduce meaning predicates, but not how those meanings should be connected to each other. Argument linking is thus performed by the grammar, and here the definite articles play a crucial role: the articles der and den, combined with singular-masculine nouns, are unambiguous cues. They can therefore be used by the transitive construction to assign nominative case to der Hund ‘the dog,’ and accusative case to den Mann ‘the man.’ In the implementation, as shown in example (4), this argument linking is achieved by making coreferential variables equal: \( ?a \) and \( ?x \) are made equal because both variables are bound to the same referent in the world (i.e. the biting dog), and \( ?b \) and \( ?y \) are made equal because both variables refer to the victim of the biting (i.e. the man). As there is only one possible interpretation left, the utterance has been disambiguated during parsing.

\[ (4) \]

Suppose now that the listener needs to parse the NHG utterance die Katze beißt die Frau ‘the cat bites the woman.’ The definite article die, which is underspecified for nominative- and accusative-singular-feminine, is not a reliable cue for disambiguating the utterance between two possible interpretations: ‘the cat bites the woman’ or ‘the woman bites the cat.’ Examples (5a) and (5b) illustrate this ambiguity through two different options to make variables equal.
Here, German speakers are likely to use other cues such as word order, intonation and world knowledge (cats are more likely to bite a person than the other way around) for disambiguating the utterance. In the experiments, however, these grammatical cues are not available to the listener, hence the utterance cannot be disambiguated during parsing.

Using this operationalization, the agents can autonomously experience whether utterances are ambiguous or not. These experiences can then be aggregated into the measures **Cue Reliability** and **Disambiguation Power**, as shown in the equations in (6) and (7). Let \( \mathcal{L}_i \) be a particular language system, and \( CR(\mathcal{L}_i) \) the cue reliability of that language system. \( CR(\mathcal{L}_i) \) is calculated by dividing the number of disambiguated utterances given the language system \( (U^D|\mathcal{L}_i) \) by the total number of utterances \( U \). A language’s disambiguation power \( DP \) is calculated in a similar way, namely by dividing the number of disambiguated utterances given all available systems in a language \( (U^D|\mathcal{L}_i, \mathcal{L}_{i+1}, \ldots, \mathcal{L}_n) \) by the total number of utterances \( U \).

\[
CR(\mathcal{L}_i) = \frac{(U^D|\mathcal{L}_i)}{U}
\]

\[
DP = \frac{(U^D|\mathcal{L}_i, \mathcal{L}_{i+1}, \ldots, \mathcal{L}_n)}{U}
\]

**Results**

Figure 2 compares the number of ambiguous utterances as experienced by the OHG (black) and NHG agents (white) when parsing and interpreting all 360 utterance types. The X-axis shows the number of utterances. The Y-axis shows four different analyses of the experiments: on the top, the number of disambiguated utterances is shown when the agents were only allowed to use the case specification of the definite articles and the number-gender information of their head nouns \( (U^D|\mathcal{L}_1) \). These results will form the basis for calculating the cue reli-
Figure 2. This chart compares Old High German (black) and New High German (white) with regard to how many utterances were disambiguated during parsing in four different analyses. In all four sets of results, the agents had access to the information carried by the determiners and the number-gender information from their head nouns ($\mathcal{L}_1$). The first set of results shows that, in isolation, the OHG determiners are more reliable for disambiguating utterances than the NHG determiners. However, as seen in the other sets of results, where the listener is also allowed to exploit other grammatical cues such as subject-verb agreement ($\mathcal{L}_2$) and/or selection restrictions ($\mathcal{L}_3$), the difference between both systems almost disappears.

The results clearly show that the OHG definite articles offer a far more reliable cue for disambiguating utterances with 232 disambiguated utterance types, as opposed to 160 using the NHG articles. Most problematic utterance types for both grammars are all types that involve a plural subject and plural direct object (compare example (8a) for OHG to example (8b) for NHG, both produced by the agents), and all types that either involve singular-neuter arguments in nominative and accusative case, or any combination of a singular-neuter and a plural argument in nominative and accusative case. For NHG, there are additional problems for disambiguating utterance types that involve singular-feminine subjects and direct objects, or any combination of singular-feminine and plural noun phrases in nominative or accusative case.
(8a) Die Man fundun deo Friuntinnā.
the.NOM/ACC.M.PL men find.PST.3PL the.NOM/ACC.F.PL (female) friends
‘The men found the (female) friends.’
or ‘The (female) friends found the men.’

(8b) Die Männer fanden die Freundinnen.
the.NOM/ACC.PL men find.PST.3PL the.NOM/ACC.PL (female) friends
‘The men found the (female) friends.’
or ‘The (female) friends found the men.’

However, when the agents are also allowed to use other cues, it becomes clear that a lot of the loss in cue reliability of the NHG system is compensated for. By also exploiting subject-verb agreement, the number of disambiguated utterance types increases from 232 to 280 for Old High German (an improvement of 13.3%) and from 160 to 256 for New High German (an improvement of 26.7%). Both grammars can profit from the fact that all types involving a nominative-singular-neuter subject with an accusative-plural object (or vice versa) are now properly disambiguated. The NHG system can additionally profit from the fact that all combinations of a nominative-singular-feminine subject with an accusative-plural object (or vice versa) are now disambiguated as well.

The third set of results shows that semantic selection restrictions have an even larger impact on the overall disambiguation power of the language. Here we see that OHG disambiguates 344 utterances (an improvement of 31.1% with respect to the individual cue reliability of the OHG articles) and NHG disambiguates 335 utterances (an improvement of 48.7% with respect to the individual cue reliability of NHG articles). Semantic selection restrictions are especially useful for disambiguating all utterance types where the verb selects for an animate subject versus an inanimate object. This is illustrated for OHG in example (9a) and for NHG in example (9b) using utterances produced by the agents. In both utterances, the definite articles provide insufficient information for disambiguating which argument is nominative and which argument is accusative. The selection restrictions of the main verb, however, offer the additional cue that is needed.

(9a) Die Man gābun dēn Friuntinnōm deo Gebā.
the.NOM.M.PL men give.PST.3PL the.DAT.F.PL (female) friends the.ACC.F.PL gifts
‘The men gave the (female) friends the gifts.’

(9b) Die Männer gaben den Freundinnen die Geschenke.
the.NOM.PL men give.PST.3PL the.DAT.PL (female) friends the.ACC.PL gifts
‘The men gave the (female) friends the gifts.’
When allowing the agents to exploit both subject-verb agreement and selection restrictions in addition to the information provided by the articles and by the number-gender information of their head nouns, we see that the difference between both grammars is reduced to three utterance types, which are due to the fact that NHG does not distinguish between nominative- and accusative-singular-feminine arguments. The OHG grammar now only fails to disambiguate 10 out of 360 utterance types (an improvement of 32.8% with respect to the first set of results), and the NHG grammar now fails to disambiguate only 13 utterance types (an improvement of 52.0%).

We can now calculate the cue reliability of both systems and the overall disambiguation power of the two grammars using the equations given in (6) and (7) above. The cue reliability of the OHG paradigm is 0.644 (i.e., speakers can rely on it for disambiguating 64.4% of the utterance types) versus only 0.444 for the NHG system (i.e., speakers can rely on it for disambiguating 44.4% of the utterance types). This difference confirms our intuition that syncretism harms cue reliability. However, the total disambiguation power of OHG is 0.972 as opposed to 0.964 for NHG, which is almost equivalent without even taking all available cues into account (e.g., noun declension, intonation and word order). So the increased syncretism and subsequent loss in cue reliability of the NHG definite articles do not seem to have harmed the language at all. This result contradicts our intuition that grammars consisting of more reliable subsystems are necessarily better at disambiguation than grammars consisting of only partially reliable subsystems. When pitched together, such systems can guarantee robust disambiguation power as long as they are sufficiently complementary to each other—a feature of language that has recently been dubbed ‘degeneracy’ in historical linguistics (Van de Velde, 2012).

A Closer Look

A remarkable observation is that, of the ten remaining ambiguities in the OHG utterances (when taking all cues into account), nine involve plural distinctions between nominative and accusative arguments. In other words, the three-way gender distinction between the plural forms die (NOM/ACC.M.PL), diu (NOM/ACC.N.PL) and deo (NOM/ACC.F.PL) did not contribute anything to the disambiguation of utterances. As it happens, all three forms have collapsed into a single definite article die in NHG, so the German language has been purged of a non-functional distinction.

Another important observation is that the dative articles in the NHG system never cause any problems of ambiguity, despite the collapse of the singular-feminine form der with the nominative-singular-masculine cell in the paradigm, and the collapse of the plural form den with the accusative-singular-masculine
cell (whose vowels were pronounced differently in OHG). The gender and number information of the head noun, and the fact that there is only one dative slot in the ditransitive construction, always suffices for disambiguation. This suggests that, whatever processes were responsible for the collapse of these articles, syncretism was allowed because no harm was done to the disambiguation power of the language.

3.2. Processing Efficiency

Processing efficiency measures the computational resources that language users need to allocate to the task of producing and parsing utterances. An intuitive prediction is that syncretic forms require more resources than forms with a transparent form-meaning mapping because syncretic forms are more ambiguous.

Measure

Processing efficiency is commonly assessed in terms of how much uncertainty arises when a listener needs to parse an utterance and when a speaker needs to verbalize a meaning. One approach is to approximate processing efficiency by measuring how much uncertainty can be reduced by exploiting probabilities (Levy, 2008). However, such frequency data are unavailable for Old High German, so we cannot employ the same approach here. Fortunately, the field of algorithm analysis has amply shown that significant conclusions about the resources required by an algorithm can be drawn without investigating its actual performance at runtime (Garey and Johnson, 1979). Applied to our case study, we can estimate how much it maximally costs for the linguistic processor to handle case, number and gender information.

Let’s consider the syncretic form die in OHG (example 10a) and NHG (example 10b). The lexical entry of that form contains a feature CASE whose value is a list of all of its possible functions:

(10a)  
\[
\text{die (OHG)}: \left[ \text{CASE}\{ \text{acc-sg-f nom-pl-m acc-pl-m} \} \right]
\]

(10b)  
\[
\text{die (NHG)}: \left[ \text{CASE}\{ \text{nom-sg-f acc-sg-f nom-pl acc-pl} \} \right]
\]

A naïve linguistic processor would need to consider three possible solutions for the OHG form and four possible solutions for the NHG form, which creates spurious ambiguity. Computational linguists have therefore developed much more powerful processing techniques. The one applied in the experiments involves paradigmatic inferencing. That is, instead of treating each form in isolation, the linguistic processor exploits information about how that form is
distinct from other forms in the same paradigm. There is solid evidence from psycholinguistics that paradigmatic inferencing indeed plays an important role in language processing (Clahsen et al., 2001). By measuring the cost of performing such inferences, we can assess the processing efficiency of a paradigm.

Roughly speaking, the linguistic processor will consider the morphological case paradigm of OHG or NHG each time it needs to handle the case-number-gender information of a determiner, noun or nominal phrase. The linguistic inventory of the agents thus includes an explicit representation of that paradigm in the form of a feature matrix, a data structure that has been shown to be adequate for handling German case (Dalrymple et al., 2009) and to be processing-friendly (van Trijp, 2011). Examples (11a) and (11b) show a schematic illustration of the abstract feature matrices for OHG and NHG.

![Old High German Feature Matrix](image)

![New High German Feature Matrix](image)

A feature matrix is simply a list of features that take complex values. As can be seen in the above examples, the case matrix of German consists of three features: NOM, ACC and DAT. The value of each feature is a list of elements. The first element isolates the dimension of CASE from the dimensions of NUMBER and GENDER, while the rest of the elements represent "feature bundles" that intertwine case, number and gender. The first element offers fast access to the
CASE value of a form for constructions for which the number or gender values of a form are irrelevant, whereas the other elements correspond to the cells of the paradigm.

For each form that contains the feature CASE, the processor goes through all feature-value pairs in the abstract feature matrix and returns an instantiated feature matrix in which all incompatible elements take the value ‘–’ and all compatible elements are left underspecified through a variable (indicated by a question mark). If only one option remains, the element is assigned the value ‘+’. Example (12) shows the instantiated feature matrix for the NHG form *die*. As can be seen, the feature matrix allows the article to be assigned nominative or accusative case, but only if it occurs with a feminine-singular or with a plural noun. All other case assignments are excluded.

<table>
<thead>
<tr>
<th>feature name</th>
<th>case</th>
<th>case-numer-gender bundles</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>?nom</td>
<td>-</td>
</tr>
<tr>
<td>ACC</td>
<td>?acc</td>
<td>-</td>
</tr>
<tr>
<td>DAT</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Similarly, the instantiated feature matrix of the NHG word form *Frauen* ‘women’ would specify that *Frauen* is compatible with any plural form, but incompatible with any singular form. When the linguistic processor then combines *die* and *Frauen* into the noun phrase *die Frauen*, it will combine the feature matrices of both forms in order to eliminate more possibilities. This combination, shown in example (13), has additionally ruled out a dative reading of *Frauen*, and has disambiguated *die* as a plural form. The resulting feature matrix is still underspecified for a nominative or accusative reading, which can be disambiguated at a later stage of processing. For instance, in a transitive utterance such as *Der Mann sah die Frauen* ‘the man saw the women,’ *der Mann* would already be disambiguated as a nominative-singular-masculine noun phrase, which entails that *die Frauen* has to be assigned accusative case.
In sum, paradigmatic inferencing requires the linguistic processor to access the paradigm, access each feature of the paradigm, and then check the feature’s name and each element in its value. The processing cost of paradigmatic inferencing can thus be measured by counting these operations (for a formal description of the linguistic processor, see De Beule, 2012). In a more formal way: let a feature matrix $FM$ consist of a set of feature-value pairs $FM = \{FV_i, FV_i+1, \ldots, FV_n\}$. Let the processing cost $PC$ of a feature matrix be the sum of the length of the matrix $|FM|$ (i.e. the amount of case distinctions) and the lengths of all feature-value pairs, multiplied by two:

$$PC(FM) = 2 \times (|FM| + \sum_{i=1}^{n} |FV_i|)$$

Given the processing cost of paradigmatic inferencing, we can extrapolate the processing efficiency of a morphological paradigm by comparing its actual cost to its maximal cost. The maximal cost of a paradigm equals the cost of a theoretical paradigm in which each cell has its own distinct marker. For German case, the most elaborate paradigm would contain 18 different forms (3 case distinctions \times 2 number distinctions \times 3 gender distinctions). By dividing the actual cost by the maximal cost, we extrapolate the processing cost of a paradigm on a scale from 0 to 1. The processing efficiency of a feature matrix $E(FM)$, which is the opposite of cost, can now simply be calculated by subtracting the interpolated cost from 1. The equation is shown in example (15).

$$E(FM) = 1 - \frac{PC(FM)}{MPC}$$

Results

With the equation given in (14), we can first calculate the maximal cost of processing an imaginary German system of definite articles that has 18 distinct forms to serve as a baseline. That cost is 54:
\[ MPC = 2 \times (|FM| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \]
\[ = 2 \times (3 + 8 + 8 + 8) \]
\[ = 54 \]

The cost of the OHG system, which only features a few collapsed cells, is 48:

\[ PC(FM_{OHG}) = 2 \times (|FM_{OHG}| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \]
\[ = 2 \times (3 + 8 + 8 + 5) \]
\[ = 48 \]

The New High German system, which has a smaller case paradigm, has a lower cost of 40:

\[ PC(FM_{NHG}) = 2 \times (|FM_{NHG}| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \]
\[ = 2 \times (3 + 6 + 6 + 5) \]
\[ = 40 \]

Applying the equation in (15), the processing efficiency of the OHG system is 0.111 compared to 0.260 for the NHG system. In other words, because the plural distinctions have collapsed into a single cell, the linguistic processor can perform paradigmatic inferencing twice as fast with the NHG system than with the OHG system—without any significant loss in disambiguation power. The results thus indicate that while both paradigms are equally good at eliminating ambiguities, the NHG system is more efficient in doing so.

3.3. Ease of Articulation

Speech is widely assumed to be a compromise between pronunciation economy on the one hand, and intelligibility on the other. In other words, speakers are assumed to prefer forms that require the least articulatory effort while at the same remaining distinct enough to be properly distinguished from other forms in the language.

Measure

A popular way to assess articulatory effort is to track the movements of articulators (such as the lips, tongue, and uvula) when pronouncing speech sounds (Perkell et al., 2002). The experiments presented in this paper do not involve a real speech system, but simulate phonological sounds using a method proposed by Stevens (2002). More specifically, each definite article construction contains a discrete representation of the phonemes required for pronouncing the article, with each phoneme described by a set of binary distinctive features (such
Table 1. This table illustrates the discrete representation of the NHG definite articles *die* and *das* in sets of distinctive features per phoneme. Irrelevant features have no value. Diphthongs are represented as two separate phonemes.

<table>
<thead>
<tr>
<th>Phonemes</th>
<th>die</th>
<th>das</th>
</tr>
</thead>
<tbody>
<tr>
<td>syllabic</td>
<td>- + - + -</td>
<td></td>
</tr>
<tr>
<td>continuant</td>
<td>- - +</td>
<td></td>
</tr>
<tr>
<td>sonorant</td>
<td>- - -</td>
<td></td>
</tr>
<tr>
<td>nasal</td>
<td>- - -</td>
<td></td>
</tr>
<tr>
<td>voice</td>
<td>+ + -</td>
<td></td>
</tr>
<tr>
<td>anterior</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>coronal</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>lateral</td>
<td>- - -</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>- + - - -</td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>- +</td>
<td></td>
</tr>
<tr>
<td>back</td>
<td>- - - + -</td>
<td></td>
</tr>
<tr>
<td>rounded</td>
<td>- -</td>
<td></td>
</tr>
<tr>
<td>long</td>
<td>+ -</td>
<td></td>
</tr>
</tbody>
</table>

As already said, each cost \( c_f(S_i, S_{i+1}) \) is measured as an edit distance, where we add the amount of deletions or insertions (i.e., the non-shared features \( F_n = \{f_1, \ldots, f_n\} = S_i \Delta S_{i+1} \)) to two times the amount of substitutions (i.e., the amount of shared features \( F_s = \{f_1', \ldots, f_m'\} = S_i \cap S_{i+1} \)) whose values are different in the two sets). For example, the articulatory effort for *die* is 14 (10 non-shared features + 4 for two shared features with a different value), whereas the effort for
This table shows the articulatory effort for each definite article on the left, and its ease of articulation on the right. The ease of pronunciation has increased from OHG to NHG through phonological reduction in the dative singular forms and by shifting the diphthongs of the nominative and accusative plural forms to a long vowel.

<table>
<thead>
<tr>
<th>Old High German</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td></td>
<td>Plural</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masc</td>
<td>Neut</td>
<td>Fem</td>
<td>Masc</td>
</tr>
<tr>
<td>NOM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.6757</td>
<td>26</td>
<td>daz</td>
</tr>
<tr>
<td>ACC</td>
<td>24</td>
<td>0.6757</td>
<td>26</td>
<td>daz</td>
</tr>
<tr>
<td>DAT</td>
<td>40</td>
<td>0.4995</td>
<td>40</td>
<td>dëmu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New High German</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td></td>
<td>Plural</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masc</td>
<td>Neut</td>
<td>Fem</td>
<td>Masc</td>
</tr>
<tr>
<td>NOM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.6757</td>
<td>26</td>
<td>das</td>
</tr>
<tr>
<td>ACC</td>
<td>24</td>
<td>0.6757</td>
<td>26</td>
<td>das</td>
</tr>
<tr>
<td>DAT</td>
<td>24</td>
<td>0.6757</td>
<td>24</td>
<td>dem</td>
</tr>
</tbody>
</table>

das is 28 (14 for moving from [d] to [a] and 14 for moving from [a] to [s]). So die is more economic than das.

Ease of articulation can then be inferred by comparing the average effort of OHG articles and NHG articles against a baseline of maximal effort. Here, I take maximal effort to be 78, which would be the effort of an imaginary article consisting of four phonemes (which corresponds to the size of the largest forms in the comparison, namely OHG dëmu and dëru) that are maximally distant from each other. Using the maximal effort, we can interpolate the articulatory effort of an article on a scale of zero to one by dividing its actual cost by the maximal effort. For instance, the interpolated effort of the NHG article die is 0.189 (= 14/78). Ease of articulation is the opposite measure of articulatory effort and is calculated by taking the difference of 1 and the interpolated effort. The ease of articulation of die is therefore 0.811.

Results

Table 2 shows two results for each article in the OHG and NHG paradigms: articulatory effort on the left, and its ease of articulation on the right. In the OHG paradigm, the average ease of articulation is 0.668 (calculated by counting the ease of articulation of each cell in the paradigm and then dividing the result by the number of cells). The NHG paradigm, on the other hand, has an ease of articulation of 0.733, so the NHG articles require on average less effort to pronounce.
Looking at the results in more detail, we see that the culprits for the lower score of OHG are the dative-singular forms dëmu and dëru, which are the most “expensive” articles as they consist of four phonemes. Interestingly, these expensive forms have undergone phonological erosion and have become more economic in NHG. Another interesting observation is that the three “cheapest” forms in the OHG paradigm (die, deo and diu), which all three end in a diphthong, have also eroded into an even more economic form die in NHG (now ending in a long vowel).

This begs the question why the other articles (surviving as der, den and das in NHG), which require more effort than die, were not further reduced by phonological erosion. The answer is semantic ambiguity. Recall from the earlier results that the three nominative and accusative plural forms in OHG did not contribute anything to the language’s disambiguation power. Likewise, the shorter dative forms in the NHG system, even though increasing the syncretism in the paradigm, were harmless for the language’s disambiguation power. However, if for instance der and den were to collapse, the NHG system would not contribute anymore to the disambiguation of nominative from accusative arguments in an utterance.

3.4. Auditory Distinctiveness

The payoff for “articulatory laziness” is that the listener needs more auditory precision in order to understand the speaker, which is measured in terms of auditory distinctiveness. That is, the more distinct a form is from other forms, the easier it is to recognize for the listener.

Measure

We can use the same distinctive feature representation for calculating the distance between an observed form and its nearest phonological neighbors. I assume here that the agents are able to recognize an article when they perceive one, so only other articles are taken into consideration for measuring auditory distinctiveness. First, the phonemes of two articles are mapped onto each other using their discrete representation as described above, which is here illustrated for the NHG articles die and das:

\[(20)\]  
\[
\begin{array}{c}
d\quad i: \\
d\quad a\quad s
\end{array}
\]
Table 3. This table shows for each article its articulatory distinctiveness with respect to its nearest phonological neighbor on the left, and the same measure interpolated on a scale from zero to one on the right.

<table>
<thead>
<tr>
<th>NOM</th>
<th>Old High German</th>
<th></th>
<th>Plural</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>Masc</td>
<td>Neut</td>
<td>Fem</td>
<td>Masc</td>
</tr>
<tr>
<td>2</td>
<td>0.0192</td>
<td>6</td>
<td>0.0577</td>
<td>6</td>
</tr>
<tr>
<td>ACC</td>
<td>2</td>
<td>0.0192</td>
<td>6</td>
<td>0.0577</td>
</tr>
<tr>
<td>DAT</td>
<td>2</td>
<td>0.0192</td>
<td>2</td>
<td>0.0192</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Old High German</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New High German</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Singular</td>
</tr>
<tr>
<td>NOM</td>
<td>Masc</td>
</tr>
<tr>
<td>2</td>
<td>0.0192</td>
</tr>
<tr>
<td>ACC</td>
<td>2</td>
</tr>
<tr>
<td>DAT</td>
<td>2</td>
</tr>
</tbody>
</table>

The total distance $D$ between two forms is calculated as the sum of all the distances between two mapped sets of distinctive features $d_f(S_i, S'_i)$:

$$D = \sum_{i=1}^{k} d_f(S_i, S'_i)$$

The distance function $d_f$ is calculated in the same way as the cost function $c_f$ described in the previous section. In our example, the distance between *die* and *das* is 18 (0 for the shared phoneme [d], 8 for the distance between [i:] and [a], and 10 for all the non-shared features between the zero pronunciation of *die* and [s] of *das*).

The articulatory distinctiveness can be extrapolated to a scale between zero and one by dividing it by the maximal distance. Here, I take the maximal distance between two articles to be 104, which would be the distance between two imaginary morphemes of four phonemes each, with a maximal distance between each phoneme.

Results

The results for each article are shown in Table 3, with the actual distance between an article and its nearest neighbor on the left, and the interpolated auditory distinctiveness on the right. The first thing to notice is that all articles have at least one very close neighbor. Even the most distinct article (*die* in the NHG system) still only scores 0.135 on a scale of zero to one. The results thus suggest
Figure 3. These spider charts each take an article at their center and then show the interpolated distances between that article and other forms of the same paradigm. The center articles are OHG *die* (top left) and *dëru* (bottom left) and their corresponding forms in NHG *die* (top right) and *der* (bottom right). When comparing the spider charts for OHG and NHG *die*, we see that a cluster of close OHG forms (*die, diu* and *deo*) have collapsed into a single form in NHG, which has improved the auditory distinctiveness of the paradigm. When comparing OHG *dëru* to NHG *der*, we see that the OHG forms of *dëmu* and *dëru* have undergone phonological reduction, which results in three close neighbors in the NHG system: *der, den* and *dem*. Despite their low auditory distinctiveness, these forms have been maintained in order to uphold the language’s disambiguation power.

that ease of articulation is more important than auditory distinctiveness, and that our auditory perception is well capable of distinguishing only minimally diverging sound patterns. On average (by adding the values of each cell and dividing the sum by the total number of cells), the NHG system scores slightly better than the OHG system, with an auditory distinctiveness of 0.077 versus 0.043.

Despite the small differences, we can still see some clusters appearing if we look at the results in more detail. Figure 3 presents four spider charts that each take one article as their center and then show the distances between that article and the other forms in its paradigm. The figure shows spider charts for the OHG articles *die* and *dëru* on the left, and two charts for their corresponding NHG forms *die* and *der* on the right.
On the top left we see the distances between OHG *die* (the center value 0) and the other OHG articles. As can be seen, the forms *diu* and *deo* are closer to the center than the other articles, which means they are harder to distinguish from each other. In the NHG system, all three forms have collapsed into the article *die*, which can be distinguished from the other NHG articles more easily, as shown in the spider chart on the top right. Given that *die*, *deo* and *diu* in OHG did not contribute to the language’s disambiguation power anyway, the loss of an unnecessary distinction has also made the task of perceiving the articles easier.

A second cluster of forms that are hard to distinguish from each other in the OHG paradigm is shown in the spider chart on the bottom left of Figure 3, which illustrates the distances between the article *dëru* in the center and the other forms in the paradigm. The chart shows that *dëmu* is the closest article, followed by *dër*. Both *dëru* and *dëmu* have undergone phonological reduction, but the result is a new cluster in NHG (shown on the bottom right) in which *der*, *den* and *dem* are very close to each other. From the viewpoint of auditory distinctiveness, this is a less than optimal arrangement, and it may seem surprising that such minimal differences have been able to survive for centuries in the German language. Here again it seems that semantic ambiguity plays the referee in deciding whether or not to uphold a case distinction: collapsing *der* and *dem* would result in ambiguities in any NOM-DAT pattern involving singular-masculine arguments, and as already said, collapsing *der* and *den* would make the system useless for distinguishing nominative from accusative arguments.

3.5. Summary

All the extrapolated measures and results are summarized in Figure 4, which compares the performance profile of the OHG paradigm (black) to that of the NHG system (white). The comparison clearly shows that the NHG system has a lower cue reliability, but that the German language has remained stable in terms of disambiguation power. In the three other measures, we see that the NHG system outperforms its OHG predecessor. The NHG articles are easier to process, pronounce and perceive than the OHG forms.

4. Discussion and Conclusion

The German language users constitute the largest speech community in Western and Central Europe with about 100 million speakers (Harbert, 2007: 16). In such a large population, which is in strong contact with other languages, it is hard to fathom how a system full of historical “accidents” manages to survive. Even harder to understand is how such accidents managed to propagate in the
first place, thereby replacing strongly entrenched forms that are more transparent in how they map meaning onto form.

This puzzle disappears in the light of the experimental results presented in this paper. Instead of being a mere by-product of various mechanical diachronic processes, the evolution of the German definite articles seems to be motivated by the communicative needs and constraints of the German language users. More specifically, the speakers of German have managed to reduce the size of their definite article system by half, going from twelve to only six forms, without harming the power of the system as a means for communication. Despite the increase in syncretism, the NHG system is more efficient for processing, pronunciation and perception, and the language as a whole remains strong at disambiguating utterances.

The details of the results also indicate that the use of clearly defined linguistic assessment criteria can greatly contribute to understanding the intricate interplay between various (often competing) linguistic pressures. For the German definite articles, for instance, we can now explain why phonological changes were able to operate more strongly on some forms than on others: as long as a language’s overall disambiguation power remains stable, linguistic forms that require less articulatory effort and are easier to perceive have a distinct advantage for linguistic
communication, and are hence more likely to propagate in a speech community. Likewise, increased syncretism does not necessarily lead to more uncertainty for the listener. On the contrary, as long as different forms are in paradigmatic opposition to each other, and as long as the language user can sufficiently rely on other cues, syncretism leads to smaller paradigms that are easier to process. These results confirm earlier studies on the definite articles in various German dialects. For instance, Shrier (1965: 436) notes that the “collapse, merger, or rearrangement of morphological items does not seem to detract from the strength of the dialect as a medium of communication. As long as there is a systematic arrangement [my stress] behind the existing distinctive inflectional suffixes, communication is not hampered among speakers of the dialect.”

The experimental method applied in this paper can be readily applied to other linguistic phenomena as well. For instance, we can implement processing models for an even older stage in the history of German, such as Proto-Germanic, and investigate why its system of demonstratives evolved into the OHG system. Such future research requires close collaboration between historical linguists, who have an indispensable grasp of the empirical evidence, and computational linguists, who can develop processing models and assessment criteria to help in the interpretation of that evidence.

References


De Beule, Joachim. 2012. A formal deconstruction of Fluid Construction Grammar. In Luc


Appendix: Linguistic Facts Relevant to the Computational Reconstruction

This Appendix offers a brief outline of the linguistic facts that matter for the computational reconstruction of OHG and NHG. An interactive web demonstration at www.fcg-net.org/demos/german-case shows how FCG parses and produces utterances.

The computational reconstructions of OHG and NHG are models of deep language processing. Both the OHG and NHG inventory consist of a basic repertoire of 24 lexical constructions (nouns and verbs) and 4 grammatical constructions (one NP construction and three argument structure constructions) that suffice for covering all 360 utterance types. On top of that, the NHG agents know six morphological constructions (one for each definite article in the language), and the OHG agents know twelve morphological constructions (one for each form in the OHG paradigm). The grammars are fully reversible, so they can be used for parsing and production.

Nouns

German nouns have two numbers (singular and plural) and three genders (masculine, feminine and neuter). Both also have several declension classes, but these differ significantly from each other. Generally speaking, OHG has a more transparent and elaborate marking of case on its nouns than NHG (Wright, 1906: Ch.IX). Consider, for instance, the noun *Tag* ‘day’ in its OHG declension versus its NHG declension (OHG example taken from Wright, 1906: 45–46), as shown in example (22).

(22) | Old High German | New High German |
---|---|---|
NOM.SG | Tag | Tag |
ACC.SG | Tag | Tag |
DAT.SG | Tage | Tag |
GEN.SG | Tages | Tag(e)s |
NOM.PL | Tagā | Tage |
ACC.PL | Tagā | Tage |
DAT.PL | Tagum | Tagen |
GEN.PL | Tago | Tage |

In order to compare the performance of the NHG system of definite articles to its OHG predecessor, however, both have to be treated on equal footing, hence such differences in noun declensions have been leveled in the implementation. For the NHG grammar, only nouns have been selected that mark a difference between singular and plural forms, but which are unmarked for case in nominative, accusative or dative slots (such as *Mann* ‘man’ vs. *Männer* ‘men’). In the
corresponding OHG nouns, case information is simply ignored. The lexical construction for each noun form thus contains information about its number and gender, but not about its case.

**Verbs and Argument Structure Constructions**

German verbs agree with the subject of an utterance in number and person. As is standard practice, verbs are associated with particular argument structure constructions by specifying their valence in their lexical entry. The computational reconstruction includes the three-place predicate *geban* (OHG) / *geben* (NHG) ‘to give,’ which is associated with the ditransitive construction (NOM-DAT-ACC pattern); and it includes two types of two-place predicates, associated with a NOM-DAT pattern (e.g., *helfan* (OHG) / *helfen* (NHG) ‘to help’) or a NOM-ACC pattern (e.g., *findan* (OHG) / *finden* (NHG) ‘to find’). Verbs may further impose semantic selection restrictions on their arguments. For instance, the beneficiary of *geben* ‘to give’ or *helfen* ‘to help’ should be [ANIMATE +].

In production, the argument structure constructions exploit the valence information of the main verb to assign cases to the verb’s arguments. In parsing, the same constructions try to exploit the case-number-gender markings on the verb’s arguments for retrieving “who did what to whom” in the utterance. Argument structure constructions do not impose any word order restrictions. Indeed, word order is ignored entirely in the experiments, as the German language uses word order as a device for marking information structure rather than for marking grammatical relations (Lenerz, 1977).

**Definite Articles**

The OHG paradigm did not consist of “definite articles” as in the NHG sense, but rather of demonstratives that could take on the function of articles (see Herbert, 2007: 142, for a discussion). As shown in example (23) (original example from Demske, 2001: 115; here quoted from Heine and Kuteva, 2006: 100), both uses already coexisted in OHG.

(23)  
\[
\text{thie árma joh thie hénti thie zeigont wóroltenti.}
\]

the arms and the hands these show world.end

‘the arms and the hands, they stand for the end of the world’

For each definite article, a morphological construction is implemented that maps a case specification onto a case marker. The case specification simply lists all the possible case-number-gender values of the article. For example, the case specification of the NHG article *dem* includes DAT-SG-M/N (dative-singular-masculine or dative-singular-neuter). The case marker is the phonological form of the definite article represented as sets of distinctive features.