Transparency versus Processing Efficiency: A Case Study on German Declension

Remi van Trijp
Sony Computer Science Laboratory Paris
6, Rue Amyot – 75005 Paris
remi@csl.sony.fr

Abstract

Ambiguity is one of the most fascinating mysteries of human language that divides the linguistic research field in roughly two camps: the mainstream view, which considers ambiguity to be undesirable for communication; and a cognitive-functional view, which argues that ambiguity allows for more efficient communication. This paper subscribes to the cognitive-functional view and presents a case study on the German declension system, which is notorious for its ambiguity through the use of syncretic case forms. Through a methodology based on computational reconstruction, the paper suggests that the current declension system outperforms its historical predecessors in terms of efficient communication, while featuring a labor-saving distribution of morphological marking across articles, adjectives and nouns.

1 Introduction

Intuitively speaking, successful communication requires languages to exhibit a transparent mapping between meaning and form. Cross-linguistic research, however, shows that such transparency is the exception rather than the rule (Leufkens, 2015). Indeed, computational linguists will sometimes jokingly admit that they only have three problems to solve: ambiguity, ambiguity, and ambiguity.[1]

Ambiguity is a major puzzle that divides the study of natural language in roughly two views (Winkler, 2015). One view, most outspokenly assumed by Chomskyan linguistics, considers ambiguity to be harmful for communication – an

[1] Also see Mannning and Schütze (1999, Chapter 1) on why ambiguity is so difficult for research on natural language processing.

assumption that is sometimes used for arguing that language is in fact "poorly designed" for communication (Chomsky, 2008, p. 136). The opposing view, defended in cognitive-functional approaches to language, argues that ambiguity makes inferential communication systems more efficient (Piantadosi et al., 2012).

This paper subscribes to the cognitive-functional approach and argues that ambiguity may lead to greater efficiency in language processing. As a case study, the paper focuses on why *case syncretism* has emerged in the German declension system. Syncretism occurs when the same form can be mapped onto different grammatical functions. One example is the English suffix *-s*, which can be used as a number marker in nouns (e.g. *cat*<u>*s*</u>) or as a conjugation marker in verbs *discover*<u>*s*</u>. Often enough, these functions may be contradictory: as a nominal marker, *-s* expresses plurality, but when used as a verbal suffix, it expresses agreement with a singular subject. The German declension system in particular is notorious for this kind of syncretism, which has puzzled many linguists for decades (see a.o. Bierwisch, 1967; Blevins, 1995; Wiese, 1996; Wunderlich, 1997; Müller, 2001; Daniels, 2001; Müller, 2002; Crysmann, 2005; Dalrymple et al., 2009).

This paper defends the hypothesis that syncretism emerged in the German declension system for efficiency reasons through a methodology that is based on computational reconstruction: it reconstructs a computational processing model of the Old High German declension system (Wright, 1906, OHG; 500–1100 AD) and of the contemporary German or New High German (NHG; from 1350 AD onwards), which can then be compared to each other in terms of communicative measures such as transparency, disambiguation power, processing efficiency, and so on. The remarkable result is that the OHG system, despite its greater degree of transparency, is less adapted for communicative efficiency than the current system. The experiments reported in this paper develops on earlier findings (van Trijp, 2011b, 2013), but investigates more closely the importance of developing adequate representations and processing models for analyzing linguistic phenomena.

2 German Declension: Not as Awful as it Seems

In his 1880 essay, the American author Mark Twain famously complained that *The awful German language* is the most "slipshod and systemless, and so slippery and elusive to grasp" language of all. The German declension system, which combines the dimensions of case, number and gender, seems to fit the bill, as it features a lot of syncretic forms. This can be seen in Table 1, which shows the German paradigm of definite articles. For instance, the definite article *der* can be used as a determiner for nouns that are (a) nominative-singular-masculine, (b) dative-singular-feminine, (c) genitive-singular-feminine and (d) genitive-plural. The language only uses six different articles for a theoretical upper bound of 24 possible combinations (4

Case	SG-M	SG-F	SG-N	PL
NOM	*der*	*die*	*das*	*die*
ACC	*den*	*die*	*das*	*die*
DAT	*dem*	*der*	*dem*	*den*
GEN	*des*	*der*	*des*	*der*

Table 1: German definite articles are marked for case, number and gender. Theoretically, the paradigm contains 24 different possible combinations (4 case distinctions × 2 number distinctions × 3 gender distinctions), but it is usually represented as containing sixteen different cells because there are no gender distinctions for plural forms. Nevertheless, the language only uses six forms for filling these sixteen cells.

case distinctions × 2 number distinctions × 3 gender distinctions). Even when discarding gender distinctions for plural nouns (leading to sixteen distinctions), each form occupies at least two cells and their corresponding functions.

2.1 Some Observations About the Declension System

Let's first review some facts about the German declension system, more particularly how it applies to articles, adjectives and nouns. German is predominantly dependent-marking, which means that most of the burden of case inflection is carried by articles and/or adjectives rather than by the head noun of an NP.

Nouns. German nouns are divided into three gender classes: masculine (e.g. *der Mann* 'the mann'), feminine (*die Frau* 'the woman'), and neuter (*das Kind* 'the child'). Plural forms are typically marked through a suffix (e.g. *die Frau-en* 'the women') and/or adding an umlaut (e.g. *die Mause* 'the mouse' vs. *die Mäuse* 'the mice'), although some nouns take zero marking (e.g. *der Löffel* 'the spoon' vs. *die Löffel* 'the spoons'). German plural marking is a complex problem in its own right (see e.g. Daelemans and Van den Bosch, 2005, Ch. 3, for experiments on the acquisition of German plurals), but for our purposes it suffices to remember that nouns usually make a distinction between singular and plural forms.

Case marking is sparse, involving only a small number of highly syncretic forms. Most feminine nouns are unmarked for case (except dative plurals, see further below). Masculine and neuter nouns usually have genitive case endings (e.g. *des Kinds* 'the child's'), unless they are so-called *n-nouns*, which take the suffix *-(e)n* for all non-nominative cases (e.g. *der Junge* 'the boy.NOM' vs. *der Junge-n* 'the

Case	SG-M	SG-F	SG-N
NOM	*ein*	*eine*	*ein*
ACC	*einen*	*eine*	*ein*
DAT	*einem*	*einer*	*einem*
GEN	*eines*	*einer*	*eines*

Table 2: German indefinite articles follow roughly the same declension pattern as the definite articles. Notable differences are the zero marking for nominative masculine and neuter forms, and the accusative neuter form. The indefinite article also has no plural form, as is the case in English.

boy-ACC/DAT/GEN'). Finally, plural nouns take the suffix *-n* in the dative case, unless they already end in *-n* or *-s* (e.g. *den Kind-er-n* 'the child-PL-DAT').

Indefinite Articles. The declension of the indefinite article *ein*, shown in Table 2, is in most cases similar to that of the definite article. The main differences are the lack of a plural form, and zero marking for nominative masculine and neuter, and accusative neuter forms.

Adjectives. German predicative adjectives are not declined for case. Attributive adjectives, on the other hand, follow three declension patterns depending on the composition of the NP in which they occur, as can be seen in Table 3. First, adjectives take strong declension if there is no preceding determiner. If, however, the adjective is preceded by a definite article (or other determiner with the same declension pattern), it takes weak declension. In this case, only two case suffixes are used: *-e* and *-en*. Finally, if the preceding determiner is indefinite, the adjective takes a mixed declension pattern.

A more general way of viewing the declension of German attributive adjectives is to say that the adjective takes on case-number-gender marking if such marking has not yet been taken by a preceding determiner. That means that if the determiner has no case suffix (e.g. *ein*), the adjective uses strong declension, as in *ein jung-er Mann* 'a young-NOM.SG.M man'.

2.2 How to Handle Case Syncretism

Many scholars have worked on the problem of the German declension system through formal grammar approaches (Bierwisch, 1967; Blevins, 1995; Wiese, 1996), and the system's syncretism has proven such a challenge (Ingria, 1990) that it

(a) Strong declension:

Case	SG-M	SG-F	SG-N	PL
NOM	*-er*	*-e*	*-es*	*-e*
ACC	*-en*	*-e*	*-es*	*-e*
DAT	*-em*	*-er*	*-em*	*-en*
GEN	*-en*	*-er*	*-en*	*-er*

(b) Weak declension:

Case	SG-M	SG-F	SG-N	PL
NOM	*-e*	*-e*	*-e*	*-en*
ACC	*-en*	*-e*	*-e*	*-en*
DAT	*-en*	*-en*	*-en*	*-en*
GEN	*-en*	*-en*	*-en*	*-en*

(c) Mixed declension:

Case	SG-M	SG-F	SG-N	PL
NOM	*-er*	*-e*	*-es*	*-en*
ACC	*-en*	*-e*	*-es*	*-en*
DAT	*-en*	*-en*	*-en*	*-en*
GEN	*-en*	*-en*	*-en*	*-en*

Table 3: Attributive adjectives take three different declension patterns depending on the presence and kind of a preceding determiner.

has inspired a lot of innovations in the apparatus employed by formal grammars (Crysmann, 2005; Daniels, 2001; Heinz and Matiasek, 1994; Müller, 1999, 2001; Sag, 2003). As is clear from the previous section, the challenge stems from the fact that each case form maps onto multiple, conflicting values that interweave three dimensions: case, number and gender.

Disjunctive Feature Representation. One traditional solution is to use *disjunctive feature representation*, in which multifunctionality is represented through listing the possibilities as disjunctions (i.e. separate alternatives). For example, the article *die* covers the nominative and accusative feminine singular case, or all plural nominative and accusative nouns. Example (1) shows a feature structure (adopted from Karttunen, 1984, p. 30) with feature-value pairs between square

brackets. Disjunctions are presented by enclosing the alternatives in curly brackets ({ }).

(1)
$$\begin{bmatrix} \text{AGREEMENT} & \left\{ \begin{bmatrix} \text{GENDER} & f \\ \text{NUM} & sg \end{bmatrix} \right. \\ & \left. \begin{bmatrix} \text{NUM} & pl \end{bmatrix} \right\} \\ \text{CASE} & \{ nom\ acc \} \end{bmatrix}$$

Despite its elegance, disjunctive feature representation is not without flaws. Crysmann (2005) argues that the grammarian is often forced to make arbitrary implementation decisions. Moreover, disjunctive features are considered to be computationally expensive representations (Flickinger, 2000). In fact, the computational complexity of unifying disjunctive features is NP-complete (Ramsay, 1990), which means grammar engineers can only reach efficient implementations by carefully managing the input and processing heuristics, or to resort to approximate results with faster algorithms (e.g. Carter, 1990; Ramsay, 1990). Some implementations have decided to eliminate disjunctions altogether whenever possible (Flickinger, 2000; Crysmann, 2005).

Figure 1 illustrates the problem. The Figure shows the search tree for parsing the utterance *Die Kinder gaben der Lehrerin die Zeichnung* 'the children gave the drawing to the (female) teacher'. The example uses a mini-grammar for German that consists of only six lexical entries: the definite articles *die* and *der*, the nouns *Kinder* 'children', *Lehrerin* 'female teacher' and *Zeichnung* 'drawing' and the verb form *gaben* 'gave.PL'. All lexical entries use disjunctive feature representation for their agreement features `case`, `gender` and `number` similar to examples (1) above. Additionally, the grammar contains a Determiner-Noun construction that imposes agreement between the determiner and its head noun, and a ditransitive construction that captures the argument structure of the utterance.

What happens is that the disjunctions cause a split in the search tree whenever there are multiple alternatives possible. For example, *die Kinder* could be nominative or accusative plural, *der Frau* could be dative or genitive singular, and *die Zeichnung* could be nominative or accusative singular. This means that the search engine potentially has to consider seven false parses before the correct one is found. Additionally, every time a branch splits, the search space balloons accordingly because the search algorithm has to consider alternative orderings of applying the same constructions. The pluses in the Figure stand for these alternative branches that lead to duplicate nodes in the search tree. Their full expansion is not shown because of space limitations, but it should be obvious that detecting and pruning such duplicate nodes is a costly matter in terms of processing effort.

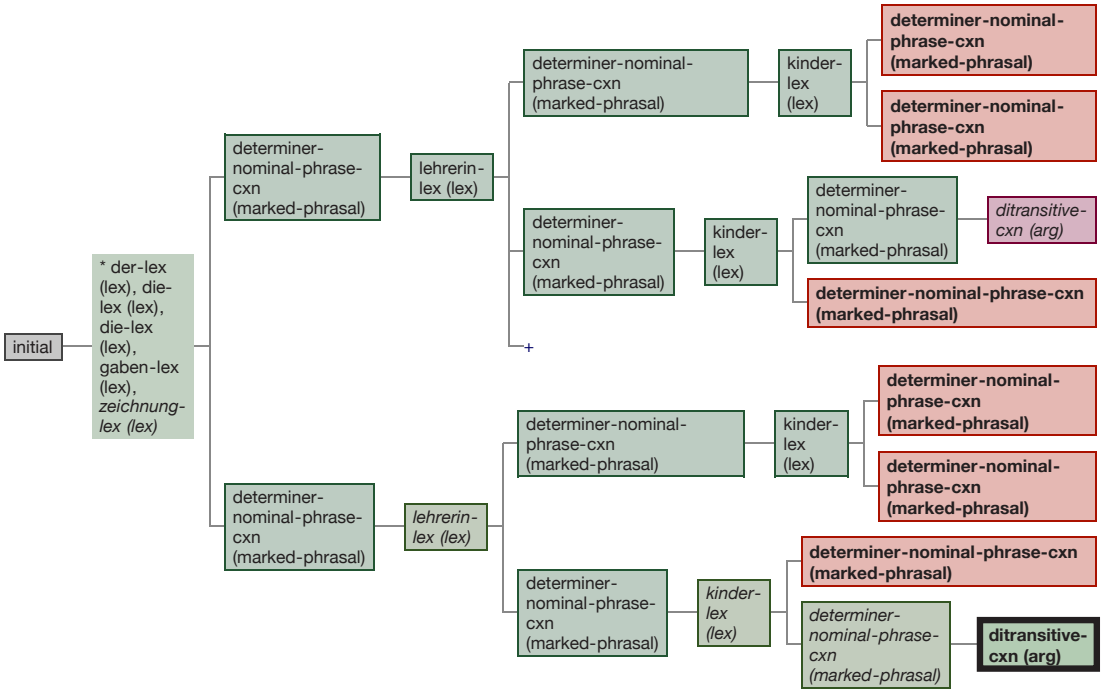

Figure 1: Parsing of the utterance *Die Kinder gaben der Lehrerin die Zeichnung* 'The children gave the drawing to the (female) teacher.' As can be seen, disjunctions force splits in the search tree regardless of syntactic context.

These efficiency issues also suggest that this search process is implausible from a psycholinguistic point of view because the example utterance is unambiguous for German speakers: *die Kinder* is the only candidate for being the subject because it is the only noun phrase that agrees with the main verb. This leaves only the accusative slot open for *die Zeichnung*, and finally, *der Lehrerin* is unambiguously assigned dative case by the verb. In other words, the search tree does not reflect the processing choices that a natural language user would make as well, and they cause ambiguities even when the syntactic context is clear for native speakers.

Type Hierarchies. Type hierarchies have been proposed as a more elegant and efficient alternative to disjunctions in most contemporary grammar formalisms. These so-called *typed* feature structure grammars classify linguistic items in terms of types, which themselves are usually organized in a multiple inheritance network. (See Figure 2) For each type, particular constraints ('type constraints') can be defined, and each type has to satisfy the type constraints of all of its supertypes

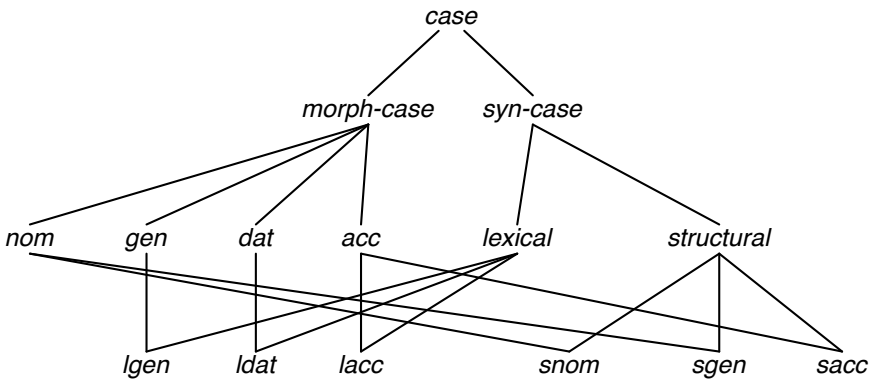

Figure 2: A type hierarchy proposed for German case agreement (Heinz and Matiasek 1994, Figure adopted from Müller 2001).

plus every constraint imposed on the type itself. A formalism's type system thus "acts as the defining framework for the rest of the grammar. For instance, it determines which structures are mutually compatible and which features can occur, and it sets up an inheritance system which allows generalizations to be expressed" (Copestake, 2002, p. 35). Even though type hierarchies do not exclude the use of disjunctions, they have sometimes been presented as a way to eliminate disjunctions whenever possible because they significantly increase efficiency (Flickinger, 2000). For German as well, various type hierarchies have been proposed (Heinz and Matiasek, 1994; Daniels, 2001; Müller, 2001).

However, the grammar of German allows for constructions that are not easily captured through a type hierarchy. Example (2) shows a coordination construction between two noun phrases. Here, *dem Mann* and the pronominal *der* share the dative case, but they differ in gender, which is not possible using a single type hierarchy for the case-number-gender combination, because structure sharing in this approach forces types to agree in number and gender as well (Müller, 2001). Solutions vary from introducing additional features (ibid.) to positing relational constraints (Daniels, 2001), but all of them return at least partially to disjunctive feature representation and therefore neutralize the efficiency gain of type hierarchies (Crysmann, 2005).

(2) *Ich helfe der und dem Mann.*
I help the.DAT.SG.F and the.DAT.SG.M man
'I help this one and the man.'

A second problem for type hierarchies is feature indetermination, as illustrated

in (3) (Pullum and Zwicky, 1986; quoted from Crysmann, 2005, p. 24):

(3) *Er findet und hilft Frauen.*
he finds.A and helps.D women.A/D
'He finds and helps women.'

The verb *finden* 'to find' normally takes an accusative complement, whereas *helfen* 'to help' takes a dative complement. *Frauen* is underspecified and can be both accusative or dative. A sentence such as **Er findet und hilft Kindern* 'He finds and helps children', on the other hand, is ungrammatical because *Kindern* can only be dative and hence clashes with the requirements of the verb *finden*. Based on such examples, it has been argued by Ingria (1990) that unification is not the best technique for syntactic agreement and case assignment, and that compatibility checks are needed instead.

Researchers have gone to great lengths to counter Ingria's claim, especially within the HPSG framework (Müller, 1999; Daniels, 2001; Sag, 2003). One solution is to augment the type hierarchy to explicitly contain neutral (or indeterminate) types (Levine et al., 2001) that can act as if they have multiple values. In example (3), the word *Frauen* would have a neutral feature so it may act as though it has both dative and accusative feature values.

Unfortunately, it is very hard to decide when types should be treated as neutral (i.e. indetermined) or ambiguous. Moreover, as argued by Crysmann (2005), such a solution leads to drastic increases in the amount of lexical ambiguity. Crysmann writes that the apparent incompatibility of feature indetermination and underspecification cannot be overcome using a single type hierarchy. Instead, he proposes two partially independent hierarchies to enable the Case values to be isolated from the Number-Gender values.

Distinctive Feature Matrices. In recent years, the puzzle of German syncretism has taken a promising turn through the use of *distinctive feature matrices* (Dalrymple et al., 2009; van Trijp, 2011b), which is an elaboration on the proposal by Ingria (1990) to no longer treat the value of a feature such as Case as atomic, but as an array of values. In this paper, I will adopt distinctive feature matrices as implemented in Fluid Construction Grammar (van Trijp, 2011b).

Returning to the example *Die Kinder gaben der Lehrerin die Zeichnung* 'the children gave the drawing to the teacher' and ignoring genitive for the time being, the `case` feature of the definite article *die* and the noun *Zeichnung* could be represented as follows:

(4) die: $\begin{bmatrix} \text{CASE} & \begin{bmatrix} \text{nom} & \textit{?nom} \\ \text{acc} & \textit{?acc} \\ \text{dat} & - \end{bmatrix} \end{bmatrix}$

(5) Zeichnung:
$$\begin{bmatrix} \text{CASE} & \begin{bmatrix} \text{nom} & ?nom \\ \text{acc} & ?acc \\ \text{dat} & ?dat \end{bmatrix} \end{bmatrix}$$

The above representation, which is a simplification for illustration purposes only, captures the fact that *die* is ambiguous for nominative and accusative through the use of variables (i.e. symbols with a question mark), but that it excludes dative (through the symbol –). *Zeichnung* can be assigned any of these three cases.

Remember from Figure 1 that disjunctive feature representation forces a split in the search tree between a nominative and accusative reading of *die Zeichnung*, even though the syntactic context is unambiguous. Feature matrices avoid this problem because they make use of underspecification. Unifying *die* and *Zeichnung* leads to the following feature matrix, which can still be assigned nominative or accusative case later on, but which already excludes dative:

(6) die Zeichnung:
$$\begin{bmatrix} \text{CASE} & \begin{bmatrix} \text{nom} & ?nom \\ \text{acc} & ?acc \\ \text{dat} & – \end{bmatrix} \end{bmatrix}$$

Let us now use this simple idea and apply it to the German declension paradigm. Each cell in the paradigm can be assigned a variable. This leads to the distinctive feature matrix for German as shown in Table 4. The first column of the table shows the dimension of case. All the other columns of the Table feature cells that represent a specific case-gender-number combination. For example, the variable `?nom-s-m` stands for 'nominative singular masculine'. Since plural forms do not mark differences in gender, only one plural cell is included for each case.

Case	**SG-M**	**SG-F**	**SG-N**	**PL**
?NOM	*?nom-s-m*	*?nom-s-f*	*?nom-s-n*	*?nom-pl*
?ACC	*?acc-s-m*	*?acc-s-f*	*?acc-s-n*	*?acc-pl*
?DAT	*?dat-s-m*	*?dat-s-f*	*?dat-s-n*	*?dat-pl*
?GEN	*?gen-s-m*	*?gen-s-f*	*?gen-s-n*	*?gen-pl*

Table 4: The feature matrix for German case.

Each linguistic item fills in as much information as possible in this case matrix. For example, the definite article *der* underspecifies its potential values and rules out all other options through '–', as shown in Table 5. Note that the variable name for the nominative case `?nom-s-m` is the same as the one for the cell of nominative-singular-masculine, which means that if the article unifies with a masculine noun, it is automatically disambiguated as a nominative article, and vice versa, if the

article is assigned nominative case, we can infer that it is masculine. The same goes for the dative case.

Case	**S-M**	**S-F**	**S-N**	**PL**
?nom-s-m	*?nom-s-m*	–	–	–
–	–	–	–	–
?dat-s-f	–	*?dat-s-f*	–	–
?gen	–	*?gen-s-f*	–	*?gen-pl*

Table 5: The feature matrix for *der*.

The string *Lehrerin* 'teacher.F.SG' rules out all plural forms but allows any case assignment. Since this noun is feminine, the single-dimension variables for case are the same ones as those that fill the singular-feminine cells in the matrix, as shown in Table 6.

Case	**S-M**	**S-F**	**S-N**	**PL**
?nom-s-f	–	*?nom-s-f*	–	–
?acc-s-f	–	*?acc-s-f*	–	–
?dat-s-f	–	*?dat-s-f*	–	–
?gen-s-f	–	*?gen-s-f*	–	–

Table 6: The feature matrix for *Lehrerin*.

Unification of *der* and *Lehrerin* only leaves the cells for dative and genitive feminine-singular open. In other words, *der Lehrerin* can only fill a dative or genitive slot. Other constructions may then later assign a '+' value to one of the two cases. The resulting feature matrix is shown in Table 7.

Case	**S-M**	**S-F**	**S-N**	**PL**
–	–	–	–	–
–	–	–	–	–
?dat-s-f	–	*?dat-s-f*	–	–
?gen-s-f	–	*?gen-s-f*	–	–

Table 7: The feature matrix for *der Leherin*.

The efficiency of this technique is illustrated in Figure 3, which shows the search tree for parsing the same utterance *Die Kinder gaben der Lehrerin die Zeichnung* using feature matrices in the grammar. As opposed to the search with disjunctions (see Figure 1), feature matrices do not cause splits in the search tree unless there is an actual ambiguity in the language. Instead, they postpone commitment to any particular value as long as possible and thus allow information and constraints to be filled in by every part of the linguistic inventory.

Besides the enormous efficiency gain and a more plausible search process, feature matrices only require unification as the standard processing mechanism without additional sources for checking compatibility of information. Moreover, as demonstrated by van Trijp (2011b), the technique is also expressive enough to deal with those cases where traditional solutions are struggling.

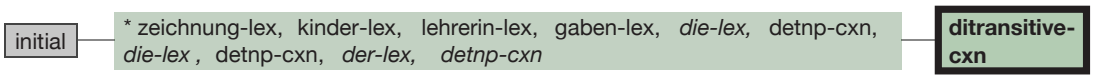

Figure 3: The search tree for *Die Kinder gaben der Lehrerin die Zeichnung* using feature matrices in the grammar.

Representing the value of the Case feature as a complex array is not only an elegant solution that rules out spurious ambiguity caused by inadequate representations, it also fits with recent psycholinguistic evidence that suggests that human language processing is made more efficient by considering paradigmatic oppositions (Clahsen et al., 2001; Eisenbeiss et al., 2005/2006). In other words, the German declension system – which had been branded as partially accidental and non-systematic – turns out to be not as awful as it seems at first sight.

3 Evaluating the Efficiency of Syncretism

Now that we have established a formal representation for case syncretism that eliminates spurious ambiguity, we need to evaluate whether syncretic forms are more or less efficient than transparent forms. This can be achieved through a novel methodology based on *computational reconstruction*. The methodology is based on the following steps (van Trijp, 2013):

1. Construct a corpus that contains relevant data for the research topic (here: constructions in which case declension matters).
2. Operationalize relevant assessment criteria. This paper uses the criteria based on language usage cue reliability, disambiguation power, processing efficiency, auditory distinctiveness, and ease of pronunciation.
3. Reconstruct a grammar sample of the language under investigation, which is capable of parsing and producing the structures found in the corpus, and evaluate its performance using the assessment criteria.
4. Replace the part of the grammar under investigation with an alternative variation and evaluate its performance using the assessment criteria.
5. Compare the results.

3.1 Experimental Set-Up

In the current experiment, we are interested in the performance of the German declension system. More specifically, we wish to test whether the system is more or less efficient than a comparable declension system that features less syncretism and more transparency in its meaning-form mappings.

The Corpus. The most important function of case marking is to identify "who does what to whom", that is, to mark the relations between a predicate and its arguments. The experiment therefore involves a corpus of declarative German utterances. There are three basic utterance *types*:

1. Ditransitive: NOM – Verb – DAT – ACC
   e.g. Die Kinder gaben der Frau die Zeichnung. ('The children gave the drawing to the woman.')

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 that the agents need to express in production 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 (Steels, 2004; van Trijp, 2011a):

(7) $\exists$ *ev,x,y: help(ev), helper(ev, x), beneficiary(ev, y), man(x), woman(y)*

All nouns in the corpus have a distinct lexical form for singular and plural (e.g. *Mann* vs. *Männer*; 'man' vs. 'men'), but are considered to be unmarked for case. That is, even if they did carry case information, this information would be ignored in the experiments. Each utterance type features subtypes that involve a unique type combination for the dimensions of number and gender, which yields 216 unique utterance subtypes for the ditransitive as follows:

(8)
	NOM.SG.M	V	DAT.SG.M	ACC.SG.M
	NOM.SG.M	V	DAT.SG.F	ACC.SG.M
	NOM.SG.M	V	DAT.SG.N	ACC.SG.M
	NOM.SG.M	V	DAT.PL.M	ACC.SG.M
			etc.	

In transitive utterances, there is an additional distinction based on animacy for noun phrases in the Object position of the utterance, which yields 72 types in the NOM-ACC configuration and 72 in the NOM-DAT configuration. Together, there are 360 unique utterance subtypes. As can be gleaned from the utterance types, the genitive case is not considered within the experiments because it is not part of basic German argument structures.

The choice for utterance *types* instead of *tokens* has been made because (a) the German language model is compared to a more transparent alternative for which no corpus data exist, and (b) the model is a so-called *precision model* that performs full analysis and production of a sentence. Precision models are currently too labor-intensive to build for achieving robust broad-coverage, hence they cannot always be used on a large corpus. However, they are very well suited for investigating the complex interplay between different linguistic constraints, so they are very good at identifying the parts of the grammar that contribute to the disambiguation task and which do not.

The use of *types* means that the model can only identify *potential* problems in language processing, whereas *tokens* would offer a more complete picture of *actual* processing complexity. It is therefore important to complement the kind of experiments as reported in this paper with models that take token frequency into account, such as the information-theoretic approaches in recent psycholinguistics studies (Hale, 2003; Levy, 2008; Piantadosi et al., 2012). However, because these models have to achieve robust broad-coverage performance, they incorporate *shallow language models* that are unable to represent all of the conditional probabilities involved in language processing, and very little work exists on language production. Future research therefore needs to find a way to combine the robustness of shallow language models with precision grammars.

The Grammar and its Variant. The experiments use a grammar sample for German implemented in Fluid Construction Grammar, capable of both parsing and producing the sentences of the corpus. Besides lexical constructions, the grammar involves a Determiner-Noun Phrase construction, and two argument structure constructions (transitive and ditransitive). Apart from a Verb-Second constraint and Subject-Verb Agreement, these argument structure constructions do not include word order information, which means that the grammar does not exploit word order for disambiguating an utterance's argument structure. Lexical verbs specify whether they take Accusative or Dative Objects, and may feature selection restrictions. The grammar uses definite articles as the sole carriers of markings of the case declension system.

In the comparative experiment, the paradigm of definite articles is replaced by a different system that features more transparency in order to test our hypothesis.

Case	SG-M	SG-F	SG-N	PL-M	PL-F	PL-N
NOM	*dër*	*diu*	*daʒ*	*die*	*deo*	*diu*
ACC	*dën*	*die*	*daʒ*	*die*	*deo*	*diu*
DAT	*dëmu*	*dëru*	*dëmu*	*dēm*	*dēm*	*dēm*
GEN	*dës*	*dëra*	*dës*	*dëro*	*dëro*	*dëro*

Table 8: The Old High German definite article paradigm.

The variant may be an artificial one, but this paper uses a reconstruction of the Old High German system (500–1100 AD; Wright, 1906), as illustrated through the OHG definite articles in Table 8. The table does not show the instrumental case, which has disappeared from the language). Choosing the OHG system is more interesting than taking an artificially constructed variant because the comparison between the current system and its actual predecessor may help us to learn something about why the OHG system evolved into its current form. One remarkable observation is that the OHG system featured twice as many forms as the current one, which begs the question why this more transparent system evolved into a more syncretic one.

3.2 Assessment Criteria and Results

The following subsections introduce the linguistic assessment criteria used in the experiments. The results are always extrapolated on a scale between zero and one, for which holds that the higher the score, the better the performance. For each measure, I first introduce an implementation-independent description, followed by its definition as operationalized in the experiments. The criteria themselves are thus assumed to be cognitively plausible, but the research does not commit itself to a specific implementation. In the remainder of this paper, I will use the term *language system* to refer to a specific grammatical aspect 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).

3.2.1 Cue Reliability and Disambiguation Power

Each language system within a particular language can be considered as a *cue* to help the disambiguation task of the listener. Consider for instance the sentences in example (9) and look at whether the Number value of the Subject of each sentence can be properly disambiguated.

(9) a. **An** antelope ran away.

b. The **man** crossed the street.

c. The fish **were** biting well that day.

d. The antelope ran away when we tried to approach **them**.

e. The antelope ran away.

In sentence (9a), the indefinite article is the cue that the speaker can rely on for disambiguating the Subject's number, whereas the noun *antelope* (which can be both singular or plural) and the verb form *ran* did not offer a reliable cue. In the next three sentences, the listener each time relies on a different cue: the singular noun form *man* (9b), the plural verb form *were* (9c), and the plural anaphoric pronoun *them* (9a). In the last sentence, the grammar does not offer the listener any help in figuring out the Number of the Subject.

Measures. *Cue reliability* is a measure that evaluates how reliable each cue (or language system) is for disambiguation without looking at other cues. The cues of example (9) all have a low reliability, because they were only helpful in one out of five utterances. The *Disambiguation Power* of a language, however, measures how often the grammar proves itself useful when all of these cues are pitched together. In example (9), the grammar was successful in four out of five cases, despite the low reliability of each individual cue. In the experiments, these measures are applied for measuring how reliable the case declension system is for disambiguating the argument structure of an utterance, and how much disambiguation power the German grammar sample has in total.

The measures are operationalized as shown in the equations in (10) and (11). 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 $(U_D|\mathcal{L}_i, \mathcal{L}_{i+1}, ..., \mathcal{L}_n)$ by the total number of utterances U.

$$CR(\mathcal{L}_i) = \frac{(U_D|\mathcal{L}_i)}{U} \tag{10}$$

$$DP = \frac{(U_D|\mathcal{L}_i, \mathcal{L}_{i+1}, ..., \mathcal{L}_n)}{U} \tag{11}$$

Results. Figure 4 compares the number of ambiguous utterances using OHG (in black) and NHG (in white) when parsing and interpreting all 360 utterance types.

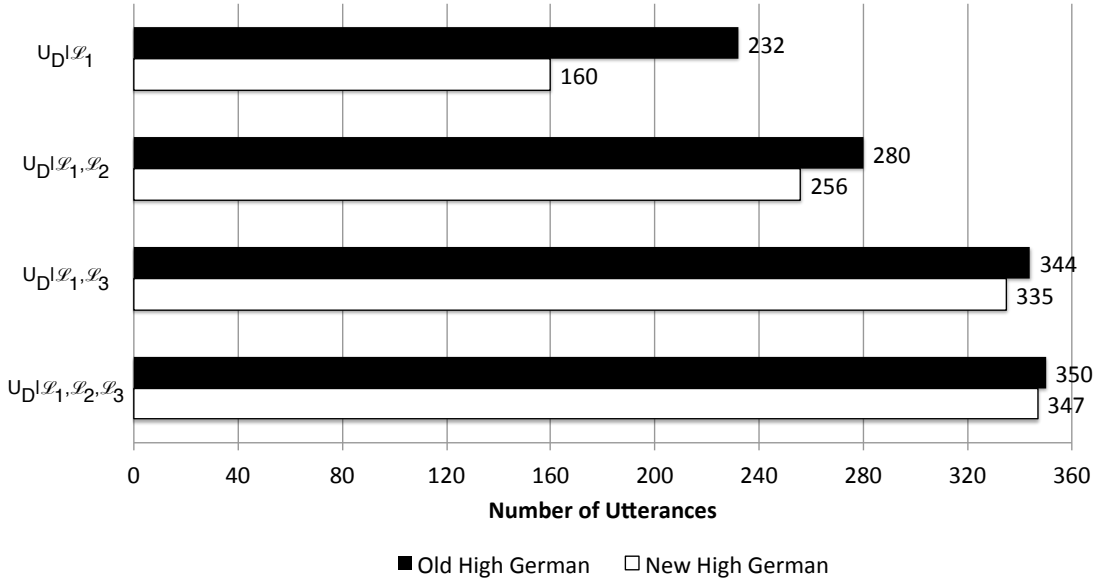

Figure 4: This chart compares Old High German (black) and New High German (white) in how many utterances were disambiguated during parsing in four different analyses. In all four set of results, the information carried by the determiners and the number-gender information from their head nouns ($\mathcal{L}_1$) were used. The first set of results show 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 the two systems almost disappears.

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 grammar only uses the case specification of the definite articles and the number-gender information of their head nouns ($U_D|\mathcal{L}_1$). These results form the basis for calculating the cue reliability of the OHG- and NHG-systems of definite articles. The other three analyses show the number of ambiguous utterances when other cues were added, such as subject-verb agreement (second set of results: $(U_D|\mathcal{L}_1, \mathcal{L}_2)$), semantic selection restrictions (third set of results: $(U_D|\mathcal{L}_1, \mathcal{L}_3)$) and both subject-verb agreement and selection restrictions (fourth set of results: $(U_D|\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$). The fourth set of results forms the basis for calculating the disambiguation power of the two systems.

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. The most problematic utterance types for both grammars are all types that involve a plural subject and plural direct object, 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.

However, when the agents are also allowed to use other cues, it becomes clear that much of the loss in cue reliability of the NHG-system can be 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 that involve 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 bigger 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 that NHG disambiguates 335 utterances (an improvement of 48.7% with respect to the individual cue reliability of NHG articles). When exploiting 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 (9-10). 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 a much smaller difference. Given the fact that the experiments did not take noun declension, intonation and word order preferences into account, the real difference may turn out to be even smaller, hence the increased syncretism in NHG definite articles does 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.

3.2.2 Processing Efficiency

Processing efficiency measures the computational resources that language users need to allocate to the task of producing and parsing utterances. We have already established that syncretic forms do not lead to bigger search spaces if some form of *paradigmatic inferencing* can be used, which is achieved in the FCG implementation through the use of distinctive feature matrices. So the question then becomes: in which case is paradigmatic inferencing more efficient?

Measure. Paradigmatic inferencing requires the linguistic processor to access each cell in a morphological paradigm. The processing cost of paradigmatic inferencing can thus be measured by counting these primitive operations. Formally: let a feature matrix FM consist of a set of feature-value pairs $FM = \{FV_i, FV_{i+1}, , FV_n\}$. 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 each feature-value pair, multiplied by two.

$$PC(FM) = 2 \times (|FM| + \sum_{i=1}^{|FM|} |FV_i|) \tag{12}$$

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 distinction $\times$ 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 (13).

$$E(FM) = 1 - \frac{PC(FM)}{MPC} \tag{13}$$

Results. With the equation given in (12), 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:

$$\begin{aligned} MPC &= 2 \times (|FM| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \\ &= 2 \times (3 + 8 + 8 + 8) \\ &= 54 \end{aligned} \tag{14}$$

The OHG-system, which only features a few collapsed cells, is 48:

$$\begin{aligned} PC(FM_{OHG}) &= 2 \times (|FM_{OHG}| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \\ &= 2 \times (3 + 8 + 8 + 5) \\ &= 48 \end{aligned} \tag{15}$$

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

$$\begin{aligned} PC(FM_{NHG}) &= 2 \times (|FM_{NHG}| + |FV_{NOM}| + |FV_{ACC}| + |FV_{DAT}|) \\ &= 2 \times (3 + 6 + 6 + 5) \\ &= 40 \end{aligned} \tag{16}$$

Applying the equation in (13), 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 at processing time, the NHG-system is more efficient in doing so.

One of the anonymous reviewers notices that the feature matrix representation for the declension system is more favourable to the NHG system than to the OHG system. While it is true that each representation is biased, the only bias of a feature matrix is to favour smaller paradigms. What is more important than this bias, is how the members of the paradigm fill each cell. If they are not strategically opposed to each other, the benefit of a small paradigm would be outweighed by the increased ambiguity. Secondly, even if the bias of a feature matrix plays a role in the results, the experiments prove that syncretism does not automatically lead to an increase in processing cost.

	die		**das**		
Phonemes	d	iː	d	a	s
syllabic	−	+	−	+	−
continuant	−		−		+
sonorant	−		−		−
nasal	−		−		−
voice	+		+		−
anterior	+		+		+
coronal	+		+		+
lateral	−		−		−
high	−	+	−	−	−
low		−		+	
back	−	−	−	+	−
rounded		−		−	
long		+		−	

Table 9: 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.

3.2.3 Ease of Articulation

One of the common assumptions about speech is that it needs to find a balance 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 measure for assessing articulatory effort is based on tracking 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 as [voice +] or [nasal –]). The distinctive feature sets used for representing the sounds of New High German articles are taken from Russ (1994); sets for Old High German are reconstructed based on descriptions by Wright (1906). Table 9 shows the distinctive feature sets for the NHG articles *die* and *das*.

Using this representation, we can calculate articulatory effort in terms of how many feature-value pairs are changed when moving from one phoneme to the next. Following a widespread tradition in computational linguistics, this paper operationalizes these changes as an *edit distance* with a few minor adaptations. Articulatory effort A is thus calculated as follows. Let the cost of moving from one set of distinctive features S_i describing a phoneme to the next one S_{i+1} be $c_f(S_i, S_{i+1})$. A is the sum of all costs until the last sound has been reached, with k as the number of phonemes, as shown in (17).

$$A = \sum_{i=1}^{k-1} c_f(S_i, S_{i+1}) \tag{17}$$

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 $F_n = \{f_1, ..., f_n\} = S_i \Delta S_{i+1}$) to two times the amount of substitutions (i.e. the amount of shared features $F_s = \{f_1, ..., f_m\} = S_i \cap S_{i+1}$ 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 *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 10 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 it 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*, who are the most

Old High German												
	Singular						Plural					
	Masc		Neut		Fem		Masc		Neut		Fem	
NOM	*dër*		*daz*		*diu*		*die*		*diu*		*deo*	
	24	0.6757	26	0.6486	20	0.7297	18	0.7568	20	0.7297	18	0.7568
ACC	*dën*		*daz*		*die*		*die*		*diu*		*deo*	
	24	0.6757	26	0.6486	18	0.7568	18	0.7568	20	0.7297	18	0.7568
DAT	*dëmu*		*dëmu*		*dëru*		*dēn*		*dēn*		*dēn*	
	40	0.4595	40	0.4595	40	0.4595	24	0.6757	24	0.6757	24	0.6757

New High German												
	Singular						Plural					
	Masc		Neut		Fem		Masc		Neut		Fem	
NOM	*der*		*das*		*die*		*die*		*die*		*die*	
	24	0.6757	26	0.6486	14	0.8108	14	0.8108	14	0.8108	14	0.8108
ACC	*den*		*das*		*die*		*die*		*die*		*die*	
	24	0.6757	26	0.6486	14	0.8108	14	0.8108	14	0.8108	14	0.8108
DAT	*dem*		*dem*		*der*		*den*		*den*		*den*	
	24	0.6757	24	0.6757	24	0.6757	24	0.6757	24	0.6757	24	0.6757

Table 10: This Table shows the articulatory effort for each definite article on the left, and its ease of articulation to the right. The ease of articulation 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.

"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 also the three "cheapest" forms in the OHG-paradigm (*die*, *deo* and *diu*), which all three end in a diphthong, have 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.

3.2.4 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. As the identification of parts-of-speech is considered to be a solved problem in computational linguistics, 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 just described above, which is here illustrated for the NHG-articles *die* and *das*:

(18)

d	iː	–
d	a	s

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

$$D = \sum_{i=1}^{k} d_f(S_i, S_{i'}) \tag{19}$$

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 [iː] 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 interpolated 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 11, 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 that ease of articulation is more important than auditory distinctiveness, and that our auditory perception is well capable of distinguishing only minimally

Old High German												
	Singular						Plural					
	Masc		Neut		Fem		Masc		Neut		Fem	
NOM	*dër*	2 0.0192	*daz̨*	6 0.0577	*diu*	6 0.0577	*die*	6 0.0577	*diu*	6 0.0577	*deo*	8 0.0769
ACC	*dën*	2 0.0192	*daz̨*	6 0.0577	*die*	6 0.0577	*die*	6 0.0577	*diu*	6 0.0577	*deo*	8 0.0769
DAT	*dëmu*	2 0.0192	*dëmu*	2 0.0192	*dëru*	2 0.0192	*dēn*	2 0.1923	*dēn*	2 0.0192	*dēn*	2 0.0192

New High German												
	Singular						Plural					
	Masc		Neut		Fem		Masc		Neut		Fem	
NOM	*der*	2 0.0192	*das*	8 0.0769	*die*	14 0.1346	*die*	14 0.1346	*die*	14 0.1346	*die*	14 0.1346
ACC	*den*	2 0.0192	*das*	8 0.0769	*die*	14 0.1346	*die*	14 0.1346	*die*	14 0.1346	*die*	14 0.1346
DAT	*dem*	2 0.0192	*dem*	2 0.0192	*der*	2 0.0192	*den*	2 0.1923	*den*	2 0.0192	*den*	2 0.0192

Table 11: 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 ten to the right.

diverging sound patterns. On average (by taking the sum of each cell and dividing it 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 5 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 anyway did not contribute to the language's disambiguation power, 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 5, which illustrates the distances between the article *dëru* in the center and the other forms

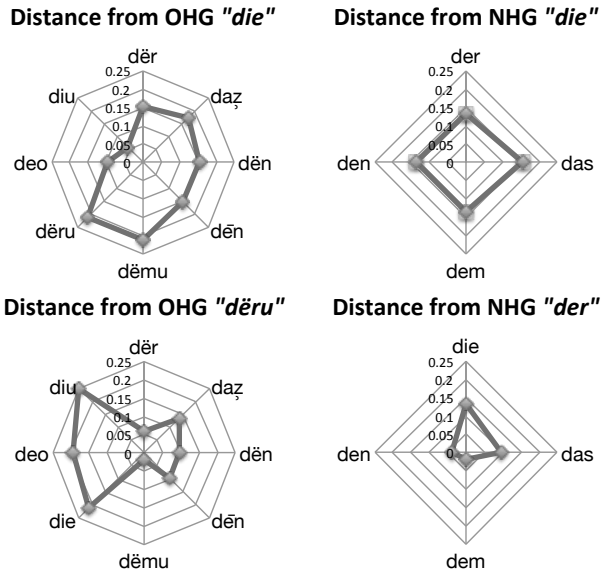

Figure 5: These spider charts each take an article at their center and then show the interpolated distances between that article and other forms of the same paradigm. The center articles are OHG *die* (top left) and *dëru* (bottom left) and their corresponding forms in NHG *die* (top right) and *der* (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.

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.

4 Discussion and Conclusions

In recent years, there has been a growing interest in the puzzle of ambiguity in natural languages. The mainstream view, most outspokenly defended by Chomskyan linguistics, ambiguity was assumed to be detrimental for communication and even as evidence that language is therefore not well-designed for serving its communicative purpose. Cognitive-functional linguists, however, claim that ambiguity serves an important communicative function. These claims are increasingly being backed up by rigorous formalization and experiments. For example, Piantadosi et al. (2012) provide information-theoretic arguments and formalization to support the claim that ambiguity makes inferential communication systems more efficient.

This paper supports this new movement through a case study on the German declension system. After reviewing the empirical observation of the system, which shows that German noun phrases typically choose either the article or the adjective for carrying case-number-gender inflection, it has proposed a formalization of the German declension system as a distinctive feature matrix in Fluid Construction Grammar. Such a feature matrix allows paradigmatic inferencing during processing, which allows case syncretisms to be efficiently processed as long as the case forms are still in functional opposition of each other. The paper then suggested that, in terms of a number of communicative assessment criteria, the current German declension system is indeed more efficient than its Old High German predecessor, which had more transparent form-meaning mappings. More specifically, ambiguity of case forms has allowed the German declension system to reduce its size in half without harming the Disambiguation Power of the language.

One important warning is that it would be wrong to conclude that the New High German system is more efficient than the Old High German one: the novel methodology proposed in this paper has plugged the Old High German system of definite articles into an otherwise New High German grammar model. In other words, the comparison only shows what the performance of the Old High German declension system would have been if it would still be part of German today. When reviewing the linguistic facts, it turns out that the Old High German definite articles only grammaticalized into articles later on. Before, they functioned as demonstratives that appeared without a head noun, which meant that they did not have access to the Gender and Number information of the head noun. In such a "linguistic ecosystem", the individual cue reliability of these case forms is of greater importance than in the current one, where there is a more established

noun phrase. In other words, it is more correct to say that the NHG declension system is more adapted to (or efficient in) the constraints of the current German grammar, whereas the OHG system seems to be more adapted to the constraints of the German grammar as it was in the OHG time period.

The result is that we get an interesting tradeoff between transparency between meaning and form on the one hand, and processing efficiency on the other. Whenever the language can rely on other linguistic cues for helping the listener's disambiguation efforts, an increase in local ambiguity can be expected. Future research will have to provide more evidence that this is indeed the case.

Acknowledgements

The research reported in this paper has been conducted at and funded by the Sony Computer Science Laboratory Paris. The author wishes to thank his colleagues for their support and constructive feedback, particularly Luc Steels and Katrien Beuls. The computational implementation for German would not have been possible without the help of my German informants Sebastian Höfer, Vanessa Micelli and Michael Spranger, and without my former colleagues Joris Bleys, Joachim De Beule, Martin Loetzsch, and Pieter Wellens, who helped to make FCG a powerful tool for exploring issues in constructional language processing. I also thank Stefan Müller for his insights on formal approaches to German, and the many researchers I interacted with when presenting this work. Special thanks also go to Aline Villavicencio and Thierry Poibeau for their patience and for offering me the chance to publish this work in a superb collection of papers, and to the reviewers for making insightful suggestions and comments. All remaining errors are of course my own.

References

M. Bierwisch. Syntactic features in morphology: General problems of so-called pronominal inflection in German. In *To Honour Roman Jakobson*, pages 239–270. Mouton De Gruyter, Berlin, 1967.

James Blevins. Syncretism and paradigmatic opposition. *Linguistics and Philosophy*, 18:113–152, 1995.

David Carter. Efficient disjunctive unification for bottom-up parsing. In *Proceedings of the 13th Conference on Computational Linguistics*, pages 70–75. ACL, 1990.

Noam Chomsky. On phases. In Robert Freidin, Carlos P. Otero, and Maria Luisa

Zubizarreta, editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pages 133–166. MIT Press, Cambridge MA, 2008.

H. Clahsen, M. Hadler, S. Eisenbeiss, and I. Sonnenstuhl-Henning. Morphological paradigms in language processing and language disorders. *Transactions of the Philological Society*, 99(2):247–277, 2001.

Ann Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, 2002.

B. Crysmann. Syncretism in german: A unified approach to underspecification, indeterminacy, and likeness of case. In S. Müller, editor, *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*, pages 91–107, Stanford, 2005. CSLI.

Walter Daelemans and Antal Van den Bosch. *Memory-Based Language Processing*. Studies in Natural Language Processing. Cambridge University Press, Cambridge, 2005.

Mary Dalrymple, Tracy Holloway King, and Louisa Sadler. Indeterminacy by underspecification. *Journal of Linguistics*, 45:31–68, 2009.

M. Daniels. On a type-based analysis of feature neutrality and the coordination of unlikes. In *Proceedings of the 8th International Conference on HPSG*, pages 137–147, Stanford, 2001. CSLI.

Sonja Eisenbeiss, Suzanne Bartke, and Harald Clahsen. Structural and lexical case in child German: Evidence from language language-impaired and typically-developing children. *Language Acquisition*, 13:3–32, 2005/2006.

D.P. Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28, 2000.

J.T. Hale. The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2):101–123, 2003.

W. Heinz and J. Matiasek. Argument structure and case assignment in German. In J. Nerbonne, K. Netter, and C. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, pages 199–236. CSLI, Stanford, 1994.

R.J.P. Ingria. The limits of unification. In *Proceedings of the 28th Annual Meeting of the ACL*, pages 194–204, 1990.

L. Karttunen. Features and values. In *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, 1984.

Sterre Leufkens. *Transparency in Language: A Typological Study*. LOT, Utrecht, 2015.

Robert Levine, Thomas Hukari, and Michael Calcagno. Parasitic gaps in english: Some overlooked cases and their theoretical consequences. In Peter W. Culicover and Paul M. Postal, editors, *Parasitic Gaps*. MIT Press, Cambridge MA, 2001.

Roger Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, 2008.

Christopher D. Mannning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge MA, 1999.

Gereon Müller. Remarks on nominal inflection in German. In Ingrid Kaufmann and Barbara Stiebels, editors, *More than Words: A Festschrift for Dieter Wunderlich*, pages 113–145. Akademie Verlag, Berlin, 2002.

Stefan Müller. An HPSG-analysis for free relative clauses in german. *Grammars*, 2(1):53–105, 1999.

Stefan Müller. Case in German – towards and HPSG analysis. In Tibor Kiss and Detmar Meurers, editors, *Constraint-Based Approaches to Germanic Syntax*. CSLI, Stanford, 2001.

J. Perkell, M. Zandipour, M. Matthies, and H. Lane. Economy of effort in different speaking conditions. i. a preliminary study of intersubject differences and modeling issues. *J. Acoust. Soc. Am.*, 112:1627–1641, 2002.

Steven T. Piantadosi, Harry Tily, and Edward Gibson. The communicative function of ambiguity in language. *Cognition*, 122:280–291, 2012.

Geoffrey Pullum and Arnold Zwicky. Phonological resolution of syntactic feature conflict. *Language*, 62(4):751–773, 1986.

Allan Ramsay. Disjunction without tears. *Computational Linguistics*, 16(3):171–174, 1990.

C. Russ. *The German Language Today. A Linguistic Introduction*. Routledge, London, 1994.

I.A. Sag. Coordination and underspecification. In J. Kom and S. Wechsler, editors, *Proceedings of the Ninth International Conference on HPSG*, Stanford, 2003. CSLI.

Luc Steels. Constructivist development of grounded construction grammars. In Walter Daelemans, editor, *Proceedings 42nd Annual Meeting of the Association for Computational Linguistics*, pages 9–19, Barcelona, 2004.

K.N. Stevens. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.*, 111:1872–1891, 2002.

Remi van Trijp. A design pattern for argument structure constructions. In Luc Steels, editor, *Design Patterns in Fluid Construction Grammar*. John Benjamins, Amsterdam, 2011a.

Remi van Trijp. A design pattern for argument structure constructions. In Luc Steels, editor, *Design Patterns in Fluid Construction Grammar*, pages 205–235. John Benjamins, Amsterdam, 2011b.

Remi van Trijp. Linguistic selection criteria for explaining language change: A case study on syncretism in German definite articles. *Language Dynamics and Change*, 3:105–132, 2013.

B. Wiese. Iconicity and syncretism. on pronominal inflection in Modern German. In R. Sckmann, editor, *Theoretical Linguistics and Grammatical Description*, pages 323–344. John Benjamins, Amsterdam, 1996.

Susanne Winkler, editor. *Ambiguity: Language and Communication*. Walter De Gruyter, Berlin, 2015.

Joseph Wright. *An Old High German Primer*. Clarendon Press, Oxford, 2nd edition, 1906.

Dieter Wunderlich. Der unterspezifizierte Artikel. In Karl Heinz Ramers Dürscheid and Monika Schwarz, editors, *Sprache im Fokus*, pages 47–55. Niemeyer, Tübingen, 1997.