ECAI 2012
Luc De Raedt et al. (Eds.)

doi:10.3233/978-1-61499-098-7-798

Self-Assessing Agents for Explaining Language Change: A Case Study in German

Remi van Trijp [1]

Abstract. Language change is increasingly recognized as one of the most crucial sources of evidence for understanding human cognition. Unfortunately, despite sophisticated methods for documenting *which* changes have taken place, the question of *why* languages evolve over time remains open for speculation. This paper presents a novel research method that addresses this issue by combining agent-based experiments with deep language processing, and demonstrates the approach through a case study on German definite articles. More specifically, two populations of autonomous agents are equipped with a model of Old High German (500–1100 AD) and Modern High German definite articles respectively, and a set of self-assessment criteria for evaluating their own linguistic performances. The experiments show that inefficiencies detected in the grammar by the Old High German agents correspond to grammatical forms that have actually undergone the most important changes in the German language. The results thus suggest that the question of language change can be reformulated as an optimization problem in which language users try to achieve their communicative goals while allocating their cognitive resources as efficiently as possible.

1 INTRODUCTION

After several decades in scientific purgatory, the study of language change has reclaimed its place as one of the most important branches in linguistics, with publications in top journals such as *Science* and *Nature* [14, 15, 18]. This renewed interest is driven by the development of quantitative methods [20, 22, 25, 42, 49] that have made it possible to reliably document the evolution of language over time. However, despite more sophisticated tools for retrieving *which* changes have taken place, the field is lacking methods for explaining *why* these changes occurred.

This paper presents a novel research method that addresses these issues by combining models of *deep* language processing (as opposed to stochastic models) with agent-based experiments. The methodology is illustrated through a case study on the German system of definite articles, whose evolution is considered to be an unsolved challenge in linguistics. The results of the case study support usage-based approaches to language [6, 26] that hypothesize that linguistic agents dynamically configure and reconfigure their language in order to satisfy their communicative needs [8, 21, 38]. From this perspective, the question of language change can be reformulated as an optimization problem in which language users try to maximize their communicative success while allocating their cognitive resources for processing and memory as efficiently as possible.

[1] Sony Computer Science Laboratory Paris, France, email: remi@csl.sony.fr

2 THE PUZZLE OF GERMAN CASE

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 language indeed seems to be full of idiosyncrasies that are hard to explain for linguists. For instance, the German definite article system (or 'paradigm') is notorious for its *syncretism* (i.e. the same form can be used for different functions), as illustrated in Table 1. 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.

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.

2.1 A HISTORICAL ACCIDENT?

Many scholars have tried to unravel the mysteries of the German definite article system through formal grammar approaches [4, 5, 50], but all of them concluded that at least part of the paradigm is non-systematic. Moreover, German case poses such hard problems for formal grammar approaches [24] that it has become the litmus test for demonstrating how well a grammar formalism copes with multifunctionality [10, 11, 23, 28, 29, 33].

The puzzle becomes all the more great when looking at the history of the German definite article paradigm. Table 2 shows how the paradigm looked like in Old High German (500–1100 AD; [51]; the table does not show the instrumental case, which has disappeared from the language). As can be seen, this system has twice as many forms, and more importantly, it has a more transparent mapping between form and function.

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

Table 2. The Old High German definite article paradigm.

Many historical linguists seem to corroborate the non-systematic analysis of formal grammarians and argue that the observed loss of transparency in the definite article paradigm is the 'accidental' by-product of sound changes and morphological processes that caused different forms to collapse [1, 2]. However, there are several problems with the 'syncretism by accident' hypothesis. First, there is abundant variation in the case systems of different German dialects and speakers are well aware of regional differences [35], so there is no explanation for why language users would prefer a non-systematic variant of the paradigm over a systematic one. Secondly, the collapsing of different case forms follows systematic tendencies instead of being randomly distributed over the paradigm [21], as would be expected if syncretism truly were a historical accident.

2.2 NOT AS AWFUL AS IT SEEMS

Recent computational studies have shed an entirely new light on German definite articles by investigating the utility of case systems for language processing. Following the hypothesis that case markers help to reduce the cognitive effort needed for interpreting utterances [36], it was found that German definite articles can speed up processing by exploiting the position that each article takes up in the whole paradigm with respect to the other ones [46], which fits with recent psycholinguistic studies on the same topic [9].

Figure 1 visualizes this processing strategy in the form of a puzzle. Suppose a listener has to parse the German utterance *Die Frau sah den Mann* ('The woman saw the man'). The form *die Frau* is 'locally' ambiguous because it can map onto nominative- or accusative-singular-feminine. In a naive formalization of German case, this local ambiguity causes additional processing effort by splitting the search tree into two different branches that need to be explored even though the utterance as a whole is unambiguous to native speakers of German. Such issues of costly search have mainly been treated as an engineering problem because many linguists assume that linguistic knowledge should be represented in a processing-independent way [34], so only few scholars have entertained the possibility that a different grammar representation might increase processing efficiency [16]. In the more efficient approach, however, the listener exploits the fact that *den Mann* only fits accusative slots in an utterance to infer that *die Frau* is nominative [46].

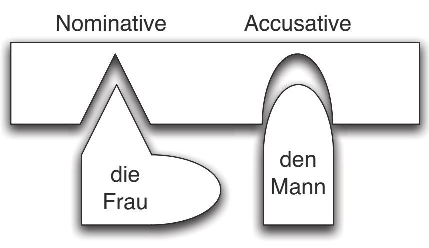


Figure 1. This Figure visualizes how grammatical knowledge can be exploited for more efficiency in processing. The rectangle represents a grammatical construction for handling utterances such as *Die Frau sah den Mann* ('The woman saw the man'). The construction has two argument slots: one that allows nominative noun phrases, and one that allows accusative phrases. A form such as *die Frau*, which is locally ambiguous because it fits both slots, can often be disambiguated because another phrase in the utterance can only occupy one slot. Indeed, *den Mann* only fits the accusative slot of the utterance, so *die Frau* can only be marked for nominative case.

Another recent study has applied this operationalization of grammar to three different variants of the German definite article paradigm – among which Old High German and Modern High German (i.e. present-day standard German) – and compared the variants in terms of processing efficiency and ambiguity resolution [48]. As it turned out, despite its syncretism, the present-day paradigm can be processed as efficiently as its Old High German predecessor without significant loss in disambiguation power. These results suggest that – rather than being a historical accident – the German case paradigm has undergone a systematic and "performance-driven [...] morphological restructuring" [21, p. 79] in which linguistic pressures such as cognitive effort decided on the maintenance or loss of certain distinctions.

2.3 SELF-ASSESSING AUTONOMOUS AGENTS

The aforementioned comparative study between Old and Modern High German argues that the evolution of the German definite article paradigm should be explained from the viewpoint of linguistic performance [48]. However, these experiments only compared the global behaviors of the different variants of the paradigm to each other. If the restructuring of grammatical systems is indeed not arbitrary but motivated by performance issues, then *individual* agents must be capable of *experiencing* inefficiencies during processing.

This paper therefore proposes to verify explanations for language change through agent-based modeling in which autonomous agents can evaluate their own linguistic behavior. This method requires linguistic assessment criteria such as communicative success, processing effort, semantic ambiguity, frequency, acoustic precision, articulatory effort, usage of short- and long-term memory, expressive adequacy, social conformity, and so on. The agents can exploit these criteria for building and maintaining *performance profiles* of the language systems they employ in communicative tasks in order to figure out which parts of their language can be optimized, very much like profiling in software engineering.

In order for the agent-based models to have explanatory scientific value, two criteria must be satisfied. First, the linguistic inefficiencies experienced by the agents need to correspond to those forms that underwent the most significant changes as attested by historical linguists. Secondly, agents are only allowed to build performance profiles based on assessment criteria that are locally observable to them. Measures such as the average inventory size, which require an unrealistic global overview of the population, are not acceptable.

3 EXPERIMENTS

The experiments feature two populations of autonomous agents with population size $N = 10$. All agents are endowed with a set of self-assessment criteria for monitoring their own linguistic behaviors. Each population is also equipped with a basic German grammar consisting of 24 lexical entries and 4 grammatical constructions. Additionally, each population has a different variant of the definite article paradigm: Old High German (which requires 12 morphological constructions) and Modern High German (which requires 6 constructions). The experiments are implemented in Babel, a multi-agent experiment framework [27]. The grammars are based on previous studies for Old and Modern High German [48] with the extension of phonological features for definite articles, and have been computationally implemented in in Fluid Construction Grammar (FCG; [37, 39]). Both FCG and Babel are available as open-source software at `www.emergent-languages.org`.

3.1 COMMUNICATIVE TASK

The agents have to produce and parse 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 [36, 45]:

(1) ∃ *ev,x,y: help(ev), helper(ev, x), beneficiary(ev, y), man(x), woman(y)*

The meanings of the arguments always have a distinct lexical form for singular and plural (e.g. *Mann* vs. *Männer*; 'man' vs. 'men'), but these are unmarked for case. All meanings are provided by a meaning generator, which ensures that the combination of arguments is always unique along the dimensions of number and gender, which yields 216 unique utterance subtypes for the ditransitive as follows:

	NOM.SG.M	V	DAT.SG.M	ACC.SG.M
	NOM.SG.M	V	DAT.SG.F	ACC.SG.M
(2)	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 by the experiments because it is not part of basic German argument structures.

One drawback of using utterance *types* instead of *tokens* is that utterance tokens offer a more reliable way of measuring *actual* performance issues, whereas *types* can only point to potential problems. However, in the current phase of the research, the use of types is warranted because there are no data available for comparing differences in token frequencies between Old and Modern High German. It would also be methodologically unacceptable to use different corpora for both variants because corpus-specific biases would distort the comparative results. Finally, the experiments feature models of *deep* language processing, so the first concern is to provide adequate coverage of the whole grammar. Obviously, however, the role of token frequency must be addressed in future research.

3.2 RESULTS

The following subsections introduce the applied self-assessment criteria and their results. The measures are operationalized using Babel's facilities for implementing *diagnostics* that can monitor every step in linguistic processing [3]. For each measure, an implementation-independent description is provided, followed by its definition as implemented in the experiments. The criteria themselves are thus assumed to be cognitively plausible, but the research does not commit itself to a specific implementation.

3.2.1 PROCESSING EFFORT

Processing effort measures the computational resources necessary for successfully producing or parsing utterances. Following one longstanding tradition in AI and psychology that models information processing in terms of *search* and strategies for exploring search spaces [17, 30], prior research has operationalized processing effort in terms of *search tree length* [41]. Using this operationalization, earlier comparisons between the definite article paradigms of Old and Modern High German have found both variants to be equivalent [48]. Here, we define processing effort in a more fine-grained way by counting the computations needed for testing which linguistic constructions can be applied and for constructing the next nodes in the search tree. In unification-based grammar formalisms such as Fluid Construction Grammar, both operations are performed by *unification* [7], so processing effort in the experiments equals the amount unifications the agents have to perform (see [12, 40] for a formal description of unification in FCG).

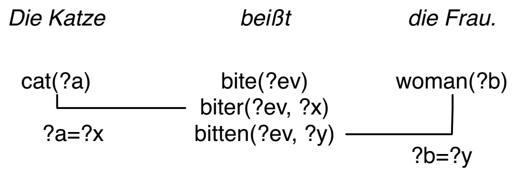
All agents had to parse and produce the 360 utterances using a standard depth-first search algorithm with each agent starting from a randomly (and differently) ordered inventory. Processing effort in the Old High German population, as self-assessed by the agents, varied between 37063173 and 37081880 unifications, whereas the Modern High German agents required only 34727788 to 34757336 unifications. Using the Welch two-sample t-test, this difference was found to be highly significant ($p = 2.2e{-}16 < 0.01$).

The results thus indicate that while both variants are equally efficient in pruning the search tree, Modern High German definite articles require less processing effort in building new search nodes. This difference is due to the fact that the Modern paradigm is only half the size of its predecessor, so the agents can extract relevant information from paradigmatic oppositions between articles more rapidly.

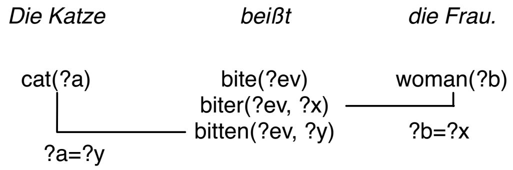
3.2.2 SEMANTIC AMBIGUITY

Semantic ambiguity equals the number of possible interpretations of an utterance. For instance, the utterance *Die Frau sah den Mann* (see Figure 1) has only one possible interpretation in which the woman is the 'seer' and the man is being 'seen'. However, an utterance such as *die Katze beißt die Frau* ('the cat bites the woman' or 'the woman bites the cat') is ambiguous because *die* has both a nominative and accusative singular-feminine reading, as illustrated in the following two examples (symbols with a question mark indicate variables):

(3) a. Interpretation 1:



b. Interpretation 2:



Here, German speakers are likely to use word order, intonation and world knowledge (i.e. cats are more likely to bite a person than the other way round) for disambiguating the utterance.

Figure 2 compares the semantic ambiguity in interpretation as experienced by an agent of each population. The left bars show the cue reliability of each paradigm for disambiguation. As can be seen, the Old High German system is significantly more reliable than the Modern High German one with 128 versus 200 ambiguities left unresolved out of 360. However, the bars on the right show the overall ambiguity when the agents also exploit other grammatical cues, such as subject-verb agreement and semantic selection restrictions. Here, the difference between the two paradigms becomes trivial, which confirms the results of earlier experiments [48].

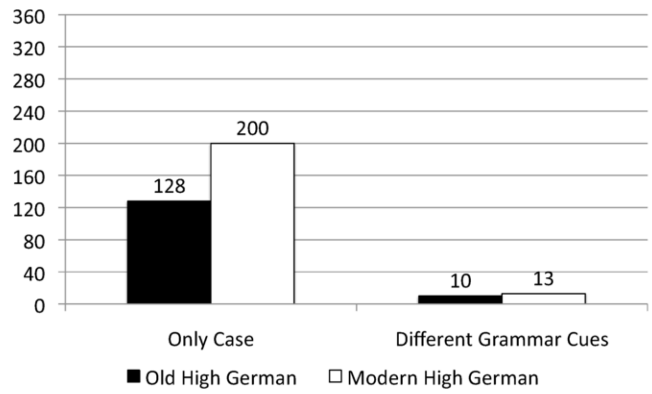


Figure 2. This Figure shows how many utterances could not be disambiguated using only information from the definite articles on the left, and the number of remaining ambiguities using all available grammar cues on the right. The results show that the Modern High German grammar is robust enough to compensate for the loss in cue reliability of its definite articles for disambiguation.

Besides monitoring overall semantic ambiguity, the agents can also keep track of which utterance types caused most semantic ambiguities. When looking at the ten remaining ambiguities using all available grammatical cues, it turns out that nine utterance types concern ambiguous plural forms (see Figure 3). More specifically, these utterances show semantic ambiguities between which phrase is nominative and accusative for the articles *die* vs. *deo* vs. *diu*. Interestingly, the three-way gender distinction for plural nominative and accusative forms has collapsed entirely in Modern High German.

3.2.3 ARTICULATORY EFFORT

Speech is widely assumed to be a compromise between pronunciation economy on the one hand, and intelligibility on the other. A popular measure for assessing *articulatory effort* is based on tracking the movements of articulators (such as the lips, tongue, and uvula) when pronouncing sounds [31].

The experiments presented in this paper do not involve a real speech system but simulate phonological sounds using a method proposed by [43]. More specifically, the lexical entries of the definite articles contain a discrete representation of the phonemes required for pronouncing the article, 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 Modern High German articles are taken from [32]; sets for Old High German are reconstructed based on descriptions by [51]. Table 3 shows the distinctive feature sets for Modern High German *die* and *das*.

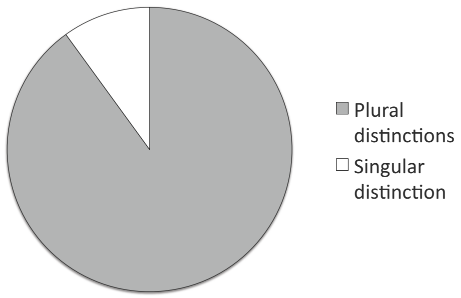


Figure 3. Nine out of ten ambiguities in the Old High German system concern plural distinctions that have collapsed in Modern High German.

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

Table 3. This Table shows the discrete representation of *die* and *das* in sets of distinctive features [32]. Irrelevant features for a phoneme have no value.

Articulatory effort A is 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:

$$A = \sum_{i=1}^{k-1} c_f(S_i, S_{i+1}) \quad (4)$$

Each cost $c_f(S_i, S_{i+1})$ is measured in a similar way as the Levenshtein distance, namely by adding the amount of 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*.

With this measure, the agents can autonomously experience which sounds require most effort. For Old High German, there are three economic articles (*die, deo* and *diu*) whose effort corresponds to 14–20. Most other articles require an articulatory effort of 24–26, except for three 'expensive' forms (*dëmu*, *dëru* and *dëro*), whose articulatory effort is 38–40. In Modern High German, those expensive forms have merged with more economic articles. However, only one of the highly economic forms (in fact, the most economic one) has been retained (*die*) and has taken over the functions of *deo* and *diu*.

3.2.4 ACOUSTIC PRECISION

The payoff for articulatory laziness is that the listener needs more *acoustic precision* in order to understand the speaker. In the experiments, the agents assess the required acoustic precision by calculating the distance between an observed form and its nearest acoustic neighbor. First, the phonemes of two articles are mapped onto each other, as illustrated as follows for *die* and *das*:

(5)

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 $d_f(S_i, S_{i'})$:

$$D = \sum_{i=1}^{k} d_f(S_i, S_{i'}) \quad (6)$$

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*).

Using this measure, the agents autonomously detected which forms are close to each other and thus require more acoustic precision. A first cluster in Old High German is *die, diu and deo*, which have a distance between 6 and 8, and which has collapsed into a single form in Modern High German. A second cluster of articles that have collapsed with other forms contains *dëmu*, *dëru* and *dëro*, which are very close to each other (distance of 2–4) but which are further away from the other articles because of their additional phoneme. Another cluster, which has largely survived in Modern High German, consists of *dër*, *dën* and *dēm* with a relative distance of 2.

4 DISCUSSION AND FUTURE WORK

Results. The experimental results confirm and move beyond earlier comparative studies on German case [48]. First, they support the hypothesis that the German definite article system has evolved to become more efficient in processing. The Old High German grammar was very successful in pruning the search tree in parsing and production, but it seems that the speakers of German found plenty of room for optimizing the system: by reducing the size of the paradigm in half, the amount of operations needed for building new nodes in the search tree has significantly dropped without losing the power to keep search effort to a minimum.

The collapsing of grammatical forms have made the system of Modern High German less reliable for semantic disambiguation. However, the results also showed that the overall semantic ambiguity in Old and Modern High German is virtually the same due to the presence of other grammatical cues such as subject-verb agreement. These results indicate that language users care less about system-specific reliability than about the performance of the intricate interactions between the language systems as a whole.

Semantic ambiguity – and the communicative success following from it – appears to have been the referee criterion in deciding which optimizations were able to propagate in the German speech community. First, the agents are able to assess the lack of a clear function for the articles *die, diu and deo* as they fail to properly disambiguate utterances. Moreover, these forms require a lot of acoustic precision on the part of the listener, hence these forms could be collapsed without harming performance. Likewise, the three most expensive forms in terms of articulatory effort (*dëmu*, *dëru* and *dëro*) have been shortened without significant increase in semantic ambiguity.

Future Work. If the German definite articles indeed evolved to become more efficient in processing, then the question of course becomes why German developed the more elaborate system in the first place. The key to answering this question lies in the fact that the Modern High German articles are optimized for the *current* 'linguistic landscape' of German. In older stages of the language (before 500 AD), the linguistic environment may well have been such that German speakers felt the communicative need to expand the system. Indeed, experiments on the origins of case systems already suggest that the same linguistic selection criteria as proposed in this paper may also drive the emergence of new grammatical forms [44, 47]. Future work therefore needs to reconstruct the predecessor of Old High German to verify this hypothesis.

Secondly, more work is needed on connecting this paper's self-assessment criteria for agents to corresponding proposals in various subfields of AI and NLP. State-of-the-art measures in each subfield are often more fine-grained than the criteria presented here, such as the use of acoustic trajectories for assessing speech economy and clarity [13], or the use of information-theoretic measures for evaluating linguistic complexity and entropy [19]. However, these more sophisticated methods are specialized for the assumptions of each domain so it is still a matter of future research to integrate them all into a model of deep language processing.

A third important future research avenue is to use agent-based experiments for explaining *how* speakers may evolve their language over time. By implementing self-assessing agents that are aware of the inefficiencies of their language systems, the first step towards this goal has been achieved. Future research therefore needs to investigate how agents can autonomously *manage* their linguistic proficiency by dynamically configuring and reconfiguring their language systems based on the performance profiles they build up through usage.

5 CONCLUSION

Language is one of the hallmarks of human cognition, and understanding how it evolves over time may help us to design more robust and open-ended artificial cognitive systems in the future. This paper showed how a combination of self-assessing agents and deep language processing can yield new explanations for language change, and it demonstrated this novel research method through a case study on the German definite article paradigm. Considered by many experts as a historical accident or a quirky idiosyncrasy, the experiments in this paper showed that the evolution of this paradigm can be explained in terms of a collective optimization process in which language users try to achieve their communicative goals while allocating their cognitive resources as efficiently as possible.

ACKNOWLEDGEMENTS

The research reported in this paper was carried out at and funded by the Sony Computer Science Laboratory Paris. This work benefited from the constructive feedback of four anonymous reviewers, as well as linguistic input from Hans Boas, Jóhanna Barðdal and Freek Van de Velde. I also wish to thank Luc Steels, director of Sony CSL, and my colleagues, for their support. All remaining errors are of course my own.

REFERENCES

[1] M. Baerman, 'Case syncretism', in *The Oxford Handbook of Case*, eds., Andrej Malchukov and Andrew Spencer, chapter 14, 219–230, OUP, Oxford, (2009).

[2] *The Role of Semantic, Pragmatic and Discourse Factors in the Development of Case*, eds., J. Barðdal and S. Chelliah, John Benjamins, Amsterdam, 2009.

[3] K. Beuls, R. van Trijp, and P. Wellens, 'Diagnostics and repairs in Fluid Construction Grammar', in *Computational Issues in Fluid Construction Grammar*, ed., L. Steels, Springer, Berlin, (2012).

[4] M. Bierwisch, 'Syntactic features in morphology: General problems of so-called pronominal inflection in German', in *To Honour Roman Jakobson*, 239–270, Mouton De Gruyter, Berlin, (1967).

[5] James Blevins, 'Syncretism and paradigmatic opposition', *Linguistics and Philosophy*, **18**, 113–152, (1995).

[6] J. Bybee and C. Beckner, 'Usage-based theory', in *The Oxford Handbook of Linguistic Analysis*, eds., B. Heine and H. Narrog, 827–855, OUP, Oxford, (2010).

[7] B. Carpenter, *The Logic of Typed Feature Structures*, CUP, Cambridge, 1992.

[8] M.H. Christiansen and N. Chater, 'Language as shaped by the brain', *Behavioral and Brain Sciences*, **31**, 489–558, (2008).

[9] 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).

[10] B. Crysmann, 'Syncretism in german: A unified approach to underspecification, indeterminacy, and likeness of case', in *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*, ed., S. Müller, pp. 91–107, Stanford, (2005). CSLI.

[11] M. Daniels, 'On a type-based analysis of feature neutrality and the coordination of unlikes', in *Proceedings of the 8th International Conference on HPSG*, pp. 137–147, Stanford, (2001). CSLI.

[12] J. De Beule, 'A formal deconstruction of Fluid Construction Grammar', in *Computational Issues in Fluid Construction Grammar*, ed., L. Steels, Springer, Berlin, (2012).

[13] B. de Boer and W. Zuidema, 'Multi-agent simulations of the evolution of combinatorial phonology', *Adaptive Behavior*, **18**(2), 141–154, (2010).

[14] M. Dunn, S.J. Greenhill, S.C. Levinson, and R.D. Gray, 'Evolved structure of language shows lineage-specific trends in word-order universals', *Nature*, **473**, 79–82, (2011).

[15] M. Dunn, A. Terrill, G. Reesink, R.A. Foley, and S.C. Levinson, 'Structural phylogenetics and the reconstruction of ancient language history', *Science*, **309**(5743), 2072–2075, (2005).

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

[17] *Heuristics. The Foundations of Adaptive Behavior*, eds., G. Gigerenzer, R. Hertwig, and T. Pachur, OUP, Oxford, 2011.

[18] R.D. Gray and Q.D. Atkinson, 'Language-tree divergence times support the anatolian theory of indo-european origin', *Nature*, **426**, 435–439, (2003).

[19] J.T. Hale, 'The information conveyed by words in sentences', *Journal of Psycholinguistic Research*, **32**(2), 101–123, (2003).

[20] D. Hall and D. Klein, 'Finding cognate groups using phylogenies', in *Proceedings of the Association for Computational Linguistics*, Uppsala, (2010). ACL.

[21] J.A. Hawkins, *Efficiency and Complexity in Grammars*, OUP, Oxford, 2004.

[22] P. Heggarty, W. Maguire, and A. McMahon, 'Splits or waves? trees or webs? how divergence measures and network analysis can unravel language histories', *Philosophical Transactions of the Royal Society*, **365**, 3829–3843, (2010).

[23] W. Heinz and J. Matiasek, 'Argument structure and case assignment in German', in *German in Head-Driven Phrase Structure Grammar*, eds., J. Nerbonne, K. Netter, and C. Pollard, 199–236, CSLI, Stanford, (1994).

[24] R.J.P. Ingria, 'The limits of unification', in *Proceedings of the 28th Annual Meeting of the ACL*, pp. 194–204, (1990).

[25] Grzegorz Kondrak, *Algorithms for Language Reconstruction*, Ph.D. dissertation, University of Toronto, 2002.

[26] R.W. Langacker, 'A dynamic usage-based model', in *Usage-Based Models of Language*, eds., M. Barlow and S. Kemmer, 1–63, Chicago University Press, Chicago, (2000).

[27] M. Loetzsch, P. Wellens, J. De Beule, J. Bleys, and R. van Trijp, 'The babel2 manual', Technical Report AI-Memo 01-08, AI-Lab VUB, Brussels, (2008).

[28] S. Müller, 'An HPSG-analysis for free relative clauses in german', *Grammars*, **2**(1), 53–105, (1999).

[29] S. Müller, 'Case in German – towards and HPSG analysis', in *Constraint-Based Approaches to Germanic Syntax*, eds., T. Kiss and D. Meurers, CSLI, Stanford, (2001).

[30] Allen Newell and Herbert A. Simon, *Human problem solving*, Prentice Hall, Englewood Cliffs, NJ, 1972.

[31] 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).

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

[33] I.A. Sag, 'Coordination and underspecification', in *Proceedings of the Ninth International Conference on HPSG*, eds., J. Kom and S. Wechsler, Stanford, (2003). CSLI.

[34] I.A. Sag and T. Wasow, 'Performance-compatible competence grammar', in *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, eds., R.D. Borsley and K. Börjars, 359–377, Wiley-Blackwell, (2011).

[35] M. Shrier, 'Case systems in German dialects', *Language*, **41**(3), 420–438, (1965).

[36] L. Steels, 'Constructivist development of grounded construction grammars', in *Proceedings 42nd Annual Meeting of the Association for Computational Linguistics*, ed., W. Daelemans, pp. 9–19, Barcelona, (2004).

[37] *Design Patterns in Fluid Construction Grammar*, ed., L. Steels, John Benjamins, Amsterdam, 2011.

[38] L. Steels, 'Modeling the cultural evolution of language', *Physics of Life Reviews*, **8**(4), 339–356, (2011).

[39] *Computational Issues in Fluid Construction Grammar*, ed., L. Steels, Springer, Berlin, 2012.

[40] L. Steels and J. De Beule, 'Unify and merge in Fluid Construction Grammar', in *Symbol Grounding and Beyond.*, eds., P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv, LNAI 4211, pp. 197–223, Berlin, (2006). Springer.

[41] L. Steels and P. Wellens, 'How grammar emerges to dampen combinatorial search in parsing', in *Symbol Grounding and Beyond. Proceedings of the Third EELC*, eds., P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv, LNAI 4211, pp. 76–88, Berlin, (2006). Springer.

[42] L. Steiner, P. Stadler, and M. Cysouw, 'A pipeline for computational historical linguistics', *Language Dynamics and Change*, **1**(1), 89–127, (2011).

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

[44] R. van Trijp, 'Grammaticalization and semantic maps: Evidence from artificial language evolution', *Linguistic Discovery*, **8**(1), 310–326, (2010).

[45] R. van Trijp, 'A design pattern for argument structure constructions', in *Design Patterns in Fluid Construction Grammar*, ed., L. Steels, John Benjamins, Amsterdam, (2011).

[46] R. van Trijp, 'Feature matrices and agreement: A case study for German case', in *Design Patterns in Fluid Construction Grammar*, ed., L. Steels, 205–236, John Benjamins, Amsterdam, (2011).

[47] R. van Trijp, 'The emergence of case marking systems for marking event structure', in *Experiments in Cultural Language Evolution*, ed., L. Steels, John Benjamins, Amsterdam, (2012).

[48] R. van Trijp, 'Not as Awful as it Seems: Explaining German Case through Computational Experiments in Fluid Construction Grammar', in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 829–839, Avignon, (2012). ACL.

[49] S. Wichmann, A. Müller, and V. Velupillai, 'Homelands of the world's language families: A quantitative approach', *Diachronica*, **27**(2), 247–276, (2010).

[50] B. Wiese, 'Iconicity and syncretism. on pronominal inflection in Modern German', in *Theoretical Linguistics and Grammatical Description*, ed., R. Sckmann, 323–344, John Benjamins, Amsterdam, (1996).

[51] J. Wright, *An Old High German Primer*, Clarendon Press, Oxford, 2nd edn., 1906.