How Grammaticalization Processes Create Grammar

Convenors:
Luc Steels (Universitat Pompeu Fabra, Barcelona)
Freek Van de Velde (University of Leuven)
Remi van Trijp (Sony Computer Science Laboratory Paris)
09:00 – 09:20  Luc Steels  
Welcome to participants and introduction to the workshop

Session 1: Finding footprints of grammaticalization  
Theme: What are novel techniques for empirically tracking grammaticalization?  
Chair: Gerhard Jaeger (University of Tübingen)

09:20 – 09:30  Xinying Chen and Tao Gong  
Evolutions of Chinese Characters ‘Zai’ and ‘Ren’ in Co-Occurrence Networks

09:30 – 09:40  Annemarie Verkerk  
Phylogenetic Comparative Methods

Session 2: Mechanisms underlying grammaticalization  
Theme: What cognitive mechanisms play a role in grammaticalization?  
Chair: Katrien Beuls (Vrije Universiteit Brussel)

2.1 Recruitment  
How can existing materials be exapted for new grammatical functions?

09:45 – 09:55  Doris Gerland  
A possible source for definiteness: possessor markers in Uralic

09:55 – 10:05  Simon Pauw and Michael Spranger  
Emergence of quantifiers: computational and robotic modeling of grammaticalization

10:05 – 10:15  Josep M. Fontana  
Changing or re-arranging? Constructional changes in perfect constructions

2.2 Analogy  
How is analogy used for expanding grammatical structure?

10:20 – 10:30  Anne Carlier and Beatrice Lamiroy  
Development of new grammatical categories: articles and auxiliaries in Romance

10:30 – 10:40  Remi van Trijp  
Analogy expands and generalizes case frames

10:40 – 10:50  Jan Nuyts  
Semantic competition and (inter)subjectification in the system of Dutch modals

2.3 Learning bias  
How do learning biases play a role in language change?

10:50 – 11:00  Robert Truswell and Nikolas Gisborne  
English WH-relatives: towards and aetiology of a gradual syntactic change

11:00 – 11:10  Jelke Bloem, Arjen Versloot and Fred Weerman  
Word order changes and grammaticalization in Germanic verbal clusters
11:15 – 11:45   Coffee break and poster session
Posters by
• Emilia Garcia Casademont
  *Interpretation process: analysis of the complexity for different language systems*
• Benjamin Fagard, Elisa Omodei and Miquel Cornudella Gaya
  *Case, prepositions, and in-betweens: sketching a model of grammatical evolution*
• Yana Knight and Michael Spranger
  *Modelling the role of Russian verbs in the evolution of Russian aspect*
• Dirk Pijpops, Katrien Beuls and Freek Van de Velde
  *Entrenchment vs. transparency. Modelling the Dutch strong-weak past tense competition in an agent-based simulation*
• Paul Van Eecke
  *Emergence and (co-)evolution of tense, aspect and modality*

Session 3: Effects from the population level
*Theme: What is the role of population structure in the propagation and convergence of languages?*
*Chair: Bernat Corominas-Murtra (Medical University of Vienna)*
11:45 – 11:55   Richard Blythe
  *Using models to relate individual linguistic behavior to the population dynamics in language*
11:55 – 12:05   Francesca Tria
  *Population structure in creole formation*
12:05 – 12:15   Peter Petré and Freek Van de Velde
  *Tracing real-life agents’ individual process in ongoing grammaticalization*

Session 4: Biological foundations
*Theme: What kind of unique capacities does the human brain have that make it ‘ready’ for language?*
*Chair: Luc Steels (Universitat Pompeu Fabra Barcelona)*
12:15 – 12:40   Eörs Szathmáry
  *Evolutionary dynamics. What is it and how is it relevant for understanding the functioning of the brain and the evolution of language?*
12:40 – 13:00   Michael Arbib
  *What does a brain need to support FCG-like capability?*

Session 5: Concluding perspectives
*Theme: What are linguistic challenges that agent-based models should address?*
*Conclusion by Salikoko Mufwene (University of Chicago) and a final discussion with all participants*
Evolutionary biology starts from observations of how species have evolved and then tries to come up with a general theory of the mechanisms that could explain how new species, and ultimately life itself, have originated. Success of evolutionary theory comes from its ability to explain concrete features of organisms, for example why cichlid fish have egg spots on their back fin, as well as general phenomena, for example why multicellular organisms have evolved. The success of evolutionary biology is undeniable. It has revolutionized all of biology and is still rolling forward at a rapid speed, aided these days by the very powerful use of computational modelling and advances in genome sequencing.

Why should Evolutionary Linguistics not strive for the same level of excellence? And how could it achieve it?

Researchers in language evolution often ignore entirely the past and ongoing change in language, despite the fact that historical linguistics has done a great job in amassing large amounts of typological and historical data and that these data prove important clues on the mechanisms that give rise to the remarkably complex structures found in human languages. Many studies of language evolution also tend to avoid using the currently available scientific methods made possible by advances in computer science: statistical analysis of language change in corpora, computer simulations, and agent-based models.

The present workshop brings together exciting work that might help to show researchers in language evolution that historical linguistics is not only relevant but indeed central and that the scientific modeling of language evolution is not only possible but in many cases highly insightful.

The workshop is organized in 5 sessions and a poster session that contains additional proposals and results. The first session (Finding footprints of grammaticalization) uses novel techniques from complex systems science to
identify grammaticalization trends. The second section (Mechanisms underlying grammaticalization) looks at various cognitive mechanisms (recruitment, analogy, learning bias) that help to explain why certain grammaticalization paths have happened. Section 3 (Effects from the population level) investigates the role of population structure and dynamics on language convergence and divergence. And section 4 examines the biological foundations of human language from the viewpoint of grammaticalization and ongoing evolution. The workshop ends with an overall perspective and concluding discussion.

1. FINDING FOOTPRINTS OF GRAMMATICALIZATION. The field of complex systems science has given rise to a whole battery of novel techniques based on analyzing the network structure of a particular phenomenon and this has been applied intensely to questions of ecology and evolution in biology. These techniques have now been applied to ongoing language change, particularly for studying the impact of linguistic context. An example of this is provided in the contribution of Chen and Gong who investigated the evolution of Chinese characters. Another battery of analysis techniques pioneered in evolutionary biology has come from the analysis of phylogenetic trees. New advances provide more sophisticated forms of analysis and they are discussed in a contribution by Verkerk.

2. MECHANISMS UNDERLYING GRAMMATICALIZATION. Agent-based modeling is a particularly appropriate method to understand the cognitive mechanisms underlying phenomena leading to the emergence of grammatical structure in human languages. Various contributions at the workshop focus on specific examples of mechanisms and combine empirical observation with attempts to create agent-based models explaining them.

2.1. RECRUITMENT. Recruitment means that there are existing forms that are used for a new purpose and then these forms begin on an evolutionary path of their own. Often the word form erodes, the meaning becomes more restricted (bleaching), and the syntactic potential more limited. An example is the evolution from demonstratives to articles. The session on recruitment discusses several case studies of this phenomenon. There are contributions on how possessor marks can become expressions of definiteness (by Gerland), how adjectives can evolve into quantifiers (by Pauw and Spranger) and how perfect constructions have evolved (by Fontana).

2.2. ANALOGY. Analogy is another cognitive mechanism that clearly plays a role in the formation of new grammatical structure. Analogy works by reorganizing certain grammatical forms so that they become similar to other ones and thus form a particular paradigm. Carlier and Lamiroy give examples of
this phenomenon for the emergence of articles and auxiliaries in Romance languages, van Trijp shows how analogy is fundamental for the structure of case paradigms, and Nuyts uses corpus data to track the evolution of meaning of the Dutch modals.

2.3. LEARNING BIAS. Many approaches to language learning assume a particular bias on the learning process, which then impacts how new grammatical structures are inductively inferred or re-arranged. The workshop shows two contributions in this direction: Truswell and Gisborne tackle change in English WH-relatives, Bloem, Versloot and Weerman look at word order change in verbal clusters.

3. EFFECTS FROM POPULATION LEVEL. Besides the cognitive mechanisms used by speakers and hearers to extend, shape and reshape their language, there is also an obvious impact of population structure, for example on which choices become dominant in a population. The work of complex systems scientists is particularly relevant for this topic and there are contributions by Blythe (who has studied the convergence of New Zealand English) and Trias (who has modeled the influence of population structure on the formation of creoles in the United States).

4. BIOLOGICAL FOUNDATIONS. Once we understand better the cognitive mechanisms that underlie the emergence of grammar, we can ask the question whether that helps us to understand the nature of the language faculty, in other words what neural mechanisms are required to support a 'language-ready' brain, that can participate in the cultural dynamics supporting language evolution. Two eminent biologists give their views on this matter: Eörs Szathmáry discusses neuronal evolutionary dynamics and Michael Arbib explores the biological bases of constructional processing.

5. CONCLUDING PERSPECTIVES. The workshop ends with a perspective by S. Mufwene and a final discussion involving all participants.
Talks
EVOLUTIONS OF CHINESE CHARACTERS ‘ZAI’ AND ‘REN’ IN CO-OCCURRENCE NETWORKS

XINYING CHEN
School of International Study, Xi’an Jiaotong University, Xianning West Road, China

TAO GONG
Department of Linguistics, University of Hong Kong, Pokfulam Road, Hong Kong

The approach of network analysis, already applied in Chinese linguistics (Li & Zhou 2007; Peng et al. 2008; Liu 2008), remains less-widely used in diachronic studies, except for one work (Liang et al. 2014). Here, we adopted this approach to study Chinese functional characters/words and evaluate their evolutions and the efficiency of this approach. We built four Chinese character co-occurrence networks based on the articles from four historical periods including ancient Chinese, middle ancient times Chinese, modern times Chinese, and modern Chinese, and then, analyzed the features of the whole networks and two characters 负 zai (a verb, meaning ‘to exist’, ‘be living’, ‘to stay or remain’; or a preposition, meaning ‘(to be located) in or at’) and 共 ren (a noun, meaning ‘human being or people’). For the sake of comparison, the chosen articles from each period had similar sizes and numbers of characters, and the two characters were frequent both in these articles and in general. In history, zai underwent a grammaticalization process, whereas ren remains as a content word. This allows us to compare the different evolutionary tendencies between the two characters.

Based on the network features (degrees, path length, density, diameter), we found that ancient Chinese underwent important changes as moving into middle ancient times, and evolved in different directions and gradually changed back, thus making modern Chinese more similar to ancient Chinese than those in the other periods. In the middle ancient times, the writing system emerged, and became more and more distinct from the oral Chinese for a long time before starting to adapt to the oral Chinese again in the modern times, and changed into modern Chinese after the reform starting in 1919 (Wang 1980). Analyses on these networks reflect such evolution of the writing system.

We also traced the evolution speeds or degrees of Chinese characters/words via these quantitative data that were rare in traditional research. We found that
the changes of the degrees of zai and ren both fit the changing tendency as described above. Two additional findings caught our attention. First, although the features of the two characters changed back after middle ancient times, the shapes of the curves were distinct. Previous research found that modern Chinese shared similar ordinate with ancient Chinese. In our study, however, the ordinates of both characters in modern Chinese were significantly higher than those in ancient Chinese, indicating that the evolution speeds or degrees of the two were higher than the average. Second, the degree of zai had a significant increase from modern times Chinese to modern Chinese, while ren had a relatively smaller increase. This could be due to the increase in the speed of the grammaticalization of zai during the time. In Solé et al.’s (2002) and Chen’s (2013) work, functional words were the hubs of either the whole network or a local community. Solé et al. (2002) also suggested that hubs could indicate the grammaticalization process and its starting points. Hubs could be functional or potential functional words to undergo future grammaticalization. Then, by analyzing the centrality of a functional word in a language network and the change of the centrality over time, we could infer the speed or degree of the grammaticalization process. Although our work showed the feasibility of using network parameters to describe the grammaticalization process, more work is needed to transform the notion of “centrality” into a numerical value for comparing words and using it to describe the evolution of functional words.

Our study showed that the network features offered a new source of information to clearly distinguish evolutions of different characters, which are relatively hard to obtain in traditional research. The network approach allowed using advanced analysis to obtain novel insights on the evolutionary tendencies of a language. This approach has enormous potentials in evolutionary research.

References


1. Introduction

The emergence of morpho-syntactical systems and the results of grammaticalization processes can be investigated through a range of different and relatively new methods, including 1) agent-based modeling (Beuls & Steels 2013), 2) corpora of historical language (Van de Velde 2010), and 3) selectionist theorizing (Steels 2011). In this paper it will be argued that there is a fourth method of interest in this respect, namely the use of phylogenetic comparative methods (Levinson and Gray 2012). Data from three different case studies are provided to illustrate the use of phylogenetic comparative methods in the study of diachronic change in morpho-syntax and grammaticalization.

2. Phylogenetic comparative methods

There are three scientific fields that undertake comparative analysis: biology, anthropology, and linguistics. Conceptual parallels between diachronic comparative studies in biology, anthropology and linguistics have been cross-fertilizing these three fields throughout history (Atkinson & Gray 2005). This is happening again with the recent adoption of phylogenetic methods from biology into linguistics. First, this was limited to the use of statistical methods for phylogenetic tree inference, i.e. analyses of how languages are related (Nichols and Warnow 2008). More recently, methods for the comparative analysis of linguistic features on the branches of a phylogenetic tree have been adopted as well (Dunn et al. 2011; Levinson & Gray 2012). These latter types of methods are called ‘phylogenetic comparative methods’ (Harvey & Pagel 1991).

These methods can be used to investigate a range of diachronic inquiries, including questions about 1) homelands of language families, 2) sequences of linguistic change, 3) dating language family trees, 4) rates of linguistic change, 5) correlations between linguistic features, and 6) ancestral states of linguistic
features (Gray et al. 2007). Whereas the methods cited in the introduction are primarily used to study change within a single language, phylogenetic comparative methods complement these by investigating change within genealogical units such as language families. How this is done and why this is relevant for the study of morphosyntactical systems and grammaticalization will be explained by means of three case studies.

3. Case studies

3.1. Indo-European motion event encoding

Motion event encoding in the Indo-European languages is extremely varied (Slobin 2004; Verkerk 2014). A range of different construction types can be used to code the same event (the following list is not exhaustive):

1. satellite-framed construction: Mary ran into the room
2. verb-framed construction: Mary entered the room (at a run)
3. deictic construction: Mary went into the room (at a run)
4. coordinate construction: Mary ran and went into the room

Germanic and Balto-Slavic languages prefer to use the first construction type; Romance languages prefer to use the second construction type; and several others, including Hindi and Armenian, prefer to use the third and fourth construction types (Verkerk 2014). Most of the attested variation is due to the merging of directional preverbs and verb roots, which affected each branch of the Indo-European family differently (Verkerk to appear). In this case study, the focus will be on how the directionality of this grammaticalization process can be incorporated into the phylogenetic analysis of construction usage.

3.2. Oceanic secondary predicates

Similar to motion event encoding in Indo-European, secondary predications of manner and result in Oceanic can take a set of different forms (Verkerk & Frostad 2013). The most important construction is the serial verb construction:

TOQABAQITA
5. Teqe kini [e qai baqita]sVC mai.
   one woman 3SG.NFUT shout be.big VENT
   ‘A woman shouted loudly this way.’ (Lichtenberk 2006:270)
In several languages, verbs from serial verb constructions have grammaticalized into adverbial particles that can no longer function as independent verbs:

**MEKEO**

6. Imi [e-biau-lobia].
   child 3SG-run-good
   ‘The child ran/has run well.’ (Jones 1998:418)

Phylogenetic comparative methods can be used to infer the behavior of Proto-Oceanic and changes in construction type along the branches of the Oceanic tree from Proto-Oceanic to the contemporary Oceanic languages (Verkerk & Frostad 2013). However, it is also possible to incorporate information on the grammaticalization of verbs from serial verb constructions into adverbial particles into the phylogenetic analysis. How this is done is shown by revisiting some of this material.

**3.3. Bantu noun-classes**

Whereas diachronic change in motion event encoding and secondary predicates has not been intensively researched, change in the Proto-Bantu noun class system has been well described (Katamba 2003, Maho 1999). However, there are still several unanswered questions: What is the nature of the relationship between noun class morphology and verbal morphology? Why are some noun classes lost more frequently than others? Can the discrepancy between SOV word order and noun class prefixes as proposed for Proto-Niger-Congo (Katamba 2003: 106-107) be unified? Answering these questions constitutes a test for the application of phylogenetic comparative methods: Are these methods able to resolve these questions and add something of importance to the study of Bantu noun classes, or does traditional comparative reconstruction suffice?

**4. Discussion**

The use of phylogenetic comparative methods complements the other approaches covered in the current workshop. They can be used alongside historical corpora, such as those used by Van de Velde (2010), to test claims about the history of individual languages against evidence from closely related languages. Studies of the emergence and loss of specific linguistic features in language families can be checked against findings about what drives these processes from agent-based models such as those used by Beuls & Steels (2013). The value of the selectionist criteria driving language change proposed
by Van Trijp (2013) and others can be assessed by looking how these criteria have interacted on the branches of phylogenetic trees of different language families. It will be demonstrated that phylogenetic comparative methods are able to incorporate information on grammaticalization in motion event encoding and secondary predication and that they have something to add even to a well-researched domain such as Bantu noun classes.

References
A POSSIBLE SOURCE FOR DEFINITENESS: POSSESSOR MARKERS IN URALIC

DORIS GERLAND
Heinrich-Heine University Duesseldorf, Universitaetsstrasse 1
Duesseldorf, 40221, Germany

The goal of this paper is firstly to show that definite artcles can evolve from possessive markers and secondly to explain why this source is a suitable starting point for the evolution of definite articles.

1. Introduction

Languages with definite articles can be found in different areas of the world and in different languages families (Dryer 2013). Most of these definite articles evolved out of demonstratives as in Indo-European languages but some of them have other sources (Himmelmann 2001). In the Samoyedic and Ob-Ugric branches of the Uralic language family, for example, possessive suffixes are applied for indicating definiteness of the host noun (Fraurud 2001, Nikolaeva 2003, Schroeder 2006). These suffixes occur in all uses regarded as typical for definite articles (Hawkins 1978), e.g. in deictic and anaphoric uses (1), in associative anaphoric uses (2) and with uniques (3). In the following referring expressions from Samoyedic languages the 3rd person singular possessive suffixes do not refer to a possessor but signal the non-ambiguity of the referent of the noun just like Indo-European definite articles do.

(1)Selkup (NOS, text2.010/2.012)
Ima [...]. Ima-timkkātii-ŋ-i-ti
woman [...] woman-3SGsay-PRS-EP-3SGOBJ
‘A woman [...] . The woman says [to him].’

(2) Nganasan (NoS, meu djamezi.003)1
stawariaa būübtaar-tu tɔʁəd’i kɔrətuu mou-n³anu s’iti ma? non’d’i-t3

1 NOS: Data of the Project “Typology of Negation in Ob-Ugric and Samoyedic languages”, University of Vienna.
now start-3SG such ordinary earth-LOC two tent stand-PRS

‘Well in the beginning [of the tale] there are two tents simply standing on
the ground.’

(3) Nganasan (Wagner-Nagy 2002:79)

Kou-ðu kantii’o čiirü’ tago
Sun-3SG disappeared cloud.PL.GEN behind

‘The sun disappeared behind the clouds.’

These uses raise different questions: What are the differences and similarities
between definite articles like those found in Indo-European languages and the
definiteness markers in Uralic languages? What are the differences and
similarities of their respective grammaticalization pathways? What licenses
possessive suffixes and especially 3rd person possessive suffixes as definite
articles? And what can we learn about definiteness in general when looking at
definite articles that emerged from a different source? Besides these questions
with regard to content the analysis of the Uralic definiteness markers raises
another important question and problem: Since we lack historical data we can
only formulate assumptions about the diachronic development and possible
pathways. So how can we find evidence for hypotheses about grammaticalization in
languages where no diachronic data is available?

2. Differences and similarities between Indo-European and Uralic
definiteness markers

The definite articles of Indo-European languages such as German, English, and
French cover a wide range of referential use. They indicate semantic
definiteness, i.e. the inherent uniqueness of the referent of the nouns they occur
with: the sun, the first man on the moon. They also indicate pragmatic
definiteness, i.e. uniqueness that comes from the context, not from the semantics
of the referent of the definiteness marked noun: the man I met yesterday, I saw a
dog [...] the dog was really big (cf. Löbner 2011 for semantic and pragmatic
uniqueness). The former uses can be regarded as reflections of the high
grammaticalization status of the definite article, its use is obligatory even though
it is redundant. The application of the definite article started with pragmatically
unique nouns in deictic and anaphoric uses and spread from there to
semantically unique nouns via analogy (cf. Demske 2001 for a case study on
German; Ortmann 2014).

The possessive suffixes of the Uralic languages do not differ from Indo-
European definite articles in their range of use but in their obligatoriness. The
analysis of different synchronic corpora shows that their application as definite article seems not to be obligatory, neither with semantic nor with pragmatically unique referents (cf. also Fraurud 2001, Nikolaeva 2003). Thus their status of grammaticalization is controversial in the literature, the non-obligatoriness speaks against a fully grammaticalized status (Lehmann 1995, Fraurud 2001) on the one hand; their occurrence with pragmatic and semantically unique referents on the other hand favours the assumption of full grammaticalization. Moreover, the Uralic possessive suffixes are still applied for indicating possession, unlike Indo-European definite articles, which differ in form and function from their original demonstrative source. However, the co-occurrence of two functions and the non-obligatoriness of a marker do not necessarily speak against a grammaticalized element. Bisang (2004) gives examples for grammaticalization without co-evolution of form and meaning and without obligatory use of the respective element. Thus it is feasible to assume that the possessive suffixes have a grammaticalized definiteness marking function as they cover all typical uses of definite articles.

Their evolution pathway might be comparable to that of definite articles in Indo-European languages if we assume general grammaticalization processes (as formulated e.g. by Hopper & Traugott 1993) without a co-evolution of form and meaning. The first step would then be the generalization and extension of the function of the element. Both demonstrative and the possessor agreement marker are anaphoric but in different ways, the former indicates the anaphoric resumption of the marked noun, the latter indicates the anaphoric resumption of the argument of the marked noun. Hence, the starting point of the grammaticalization of possessive suffixes is not “plain” anaphora as for demonstratives but associative anaphora (Fraurud 2001). From there the use might be extended and the original function might be bleached out in different ways, respectively. Data from Finnish dialects suggest that first the number specification of the possessor suffix was lost und later the person specification (see below). In a second step the source function of the element is bleached out. Demonstratives lose their primary deictic function as they became definite articles (Himmelmann 2001), possessive suffixes in definiteness marking lose their primary possessor agreement function. However, for both elements the original function is still available; most Indo-European languages allow definite articles in demonstrative function, the co-occurrence of both functions in Uralic is described above.
3. What licenses 3rd person possessor suffixes as definite articles?

With demonstratives it is mostly the distal form that is the source of definite articles (Himmelmann 1997). The application of the 3rd person possessor suffixes as definiteness marker is comparable since this form can also be conceived of as the most distal among the singular person markers (cf. the person scales proposed e.g. by Comrie 1981); the plural markers are ineligible because of their non-unique reference. However, the two main reasons why the 3rd person possessor suffix is qualified for definiteness marking are the following: (i) As a possessive pronoun the suffix originally refers to an already established and unique entity, it functions as anaphor and indicates both possession and definiteness (like associative anaphor does: My car is old. Its engine is broken.). Marked with this suffix the whole NP is definite, too. Therefore the marked head noun is interpreted as unique. In this sense the suffix marks uniquely referring expression like definite articles in other languages do. (ii) The original and still remaining function of the suffix is to indicate a possessor argument. This does not apply in cases where no relation of possession is available; with uniquely referring expressions the function of indicating a possessor is lost completely. This way, the use of the suffix is extended to contexts without a possessor, and what applies is the common denominator of indicating (either semantic or pragmatic) uniqueness.

The use of 3rd person possessor suffixes in some dialects of Finnish and Estonian can be considered as reflection of the intermediate steps of the grammaticalization pathway. In these Uralic languages the suffix is not used as a definite article but as a kind of default possessor marker (Toivonen 1998), irrespectively of the person of the possessor, as is illustrated in (4).

(4) South-West Finnish (Toivonen 1998:44)

_No tääl läksi sinä vielä asut emäntine-nsä?
well here.Q you.SG still live wife-3[SG]
‘So, do you still live here with your wife?’_

The person specification is already bleached out but the indication of an argument is still present. Note, that according to Toivonen (1998) the number specification is completely lost in South-Western Finnish, thus the suffix can also be used with plural possessors. This kind of application might display how the use of the possessive suffixes was extended.

In languages where the possessive suffixes are used both as possessor agreement marker and as definite article, the core function of the suffix seems to be to link two entities. In a possessive construction the suffix indicates the possessor and the link between possessor and possessum. In a non-possessive
construction the indication of any possessor is not relevant. What remains is the function of establishing a relation, either to the discourse situation (with pragmatically unique referents) or to cultural knowledge (with semantically unique referents). This way the definite articles with possessive sources function in the same fashion as definite articles with demonstratives sources do.

Acknowledgements

The research for this article was supported by the DFG (Deutsche Forschungsgemeinschaft) grant Collaborative Research Centre 991: “The Structure of Representations in Language, Cognition, and Science”. I would like to thank Christian Horn and Albert Ortmann for discussions.

References


EMERGENCE OF QUANTIFIERS - COMPUTATIONAL AND ROBOTIC MODELING OF GRAMMATICALIZATION

SIMON PAUW
Aldebaran A-Lab, 168 bis 170 rue Raymond Losserand
Paris, 75014, France

MICHAEL SPRANGER
Sony Computer Science Laboratories Tokyo, 3-14-13, Higashigotanda, Shinagawa-ku, Tokyo, 141-0022, Japan

The semantics of gradable quantifiers (such as “many” and “few”) depends on a number of factors ranging from size and density to expected norms. Importantly, gradable quantifiers play an important role in alleviating effects of perceptual deviation (Spranger & Pauw, 2012) typically occurring in grounded language use scenarios such as interactions about the real world. Interlocutors in such interactions often have different perspectives on the scene and therefore divergently perceive and estimate the objects and their properties in the world. In such cases, a graded notion of quantifiers that allows for lenient interpretation, i.e. margins of deviation, becomes an important communicative tool.

In the first part of the talk, we will present a computational investigation that argues how to model the semantics of quantifiers using a novel vagueness algorithm. The computational model is compared to traditional type-theory based models. We show that our notion of vagueness performs better with respect to success in communication then traditional approaches (Pauw & Spranger, 2012). The main argument from this line of research is that given the right notion of vagueness, quantifiers can be cognitively efficient and successful communication tools (Pauw & Hilfery, 2012).

Starting from the modelling of the semantics of quantifiers, we then go on to explore the emergence and grammaticalisation of graded quantifiers. The first line of experiments builds directly on the result of the semantic modelling and shows that the efficiency of graded quantifiers can be a driving force for their emergence (Pauw & Hilfery, 2012). A second line of experiments goes on to trace the dual nature of graded quantifiers. Quantifiers such as “few” and “many” can be used both as adjectives and quantifiers. Historically speaking they started out as adjectives and later took on additional functions (Solt, 2009).
the idea that the *cognitive overlap* (Durgin, 1995) of “few” and “many” with the adjectives “big” and “small” can account for their adjectival use and how *cognitive effort* and the interaction with existing quantifiers can give rise to historically attested grammaticalisation trajectories (Pauw, 2013).

**References**


One of the diachronic syntactic phenomena that has been most widely studied is the emergence of the so called *perfect tense* in modern Romance and Germanic languages. While they differ on essential points such as the gradualness or the abruptness of the change, proponents of the traditional grammaticalization or reanalysis approaches assume that at some point what was a lexical verb with a possessive meaning equivalent to contemporary English HAVE in (1a) was transformed into a fundamentally different type of expression, the functional category known as perfect tense auxiliary HAVE. Implicitly or explicitly most authors also consider that there was also a change in the category of the participle which went from being an adjective to becoming a verbal participle. Thus the construction illustrated in (1b), often referred to as a *resultative* and sometimes also as a *stative* construction, forms a minimal pair with the the perfect construction in (1c) and they are assumed to illustrate the grammatical reanalysis that gave rise to the modern perfect construction.

(1)  
\begin{align*}
\text{a. I have a nice home} \\
\text{b. Ic hæfde hine gebundenne} \\
\text{I had him tied (from Denison 1993)} \\
\text{c. I have tied him}
\end{align*}

In this talk I will discuss diachronic data from Latin and medieval Romance languages such as Old Catalan or Old Spanish together with theoretical constructs from contemporary work on lexical semantics to argue that in fact there has been no such emergence and that the way in which modern Romance languages Germanic languages express the perfect tense does not radically differ from the way in which Latin and probably also the precursors of medieval Germanic languages expressed this same tense.

The main goal of the presentation will be to show that the diachronic data involving constructions with HAVE or BE and a participle strongly support the
A discussion of the contexts in which examples like those in (2) appear will show that already in classical Latin the verb HAVE did not have a purely possessive meaning and participated together with BE in the creation of two basic types of construction that were used to express tense/aspect values identical to those expressed with the so-called HAVE perfect. The examples in (3) and (4) from Old Catalan and Old Spanish illustrate the use of other auxiliary-like verbs (TENER ("have") and ESTAR ("be") in contexts where they can be shown to convey the same type of semantic interpretations as constructions involving HAVE and BE. Crucially, these interpretations are not possible with constructions involving these verbs in modern Catalan and Spanish.

After examining the qualitative data I will present the results of a quantitative study conducted in Sánchez-Marco (2012).

As the discussion of data such as that displayed in Figure 1 below will show, after a period in which the different constructions involving participles and the verbs HAVER/HABER, SER, ESTAR and TENIR/TENER (see examples in 4 below) competed for some of the same interpretations they gradually became specialized to convey the meanings they have today. As is well known, SER appears in the so-called passive constructions as well as the copula in predicative constructions involving adjectives (participles as well as non-deverbal...
adjectives); ESTAR is also found in predicative constructions argued to have the interpretation of stage level predicates and TENER is found in constructions that given their interpretation should be considered resultative or stative constructions as the one in (1a). I will use basic standard linguistic argumentation to refute current standard analyses of some of these constructions and show that there is no reason why we should not conclude that all of them, including the so-called passive constructions, are also perfect tense constructions. In essence I will argue that, when they are studied carefully, what the patterns of evolution of these constructions suggest is not a radical reanalysis of HAVE and BE as most authors have suggested until now but rather a case of related constructional changes, crucially not involving any change in the grammatical categories of the constructions involved. I will discuss the data from some Spanish dialects and French as well as some Germanic languages where HAVE + participle constructions have come to convey the perfective aspect to determine whether in these cases we could talk about cases of constructionalization where some type of reanalysis of the grammatical categories has taken place.

Figure 5.1: Frequency of haber (top left), ser (top right), tener (bottom left) and estar (bottom right) plus aspectual classes of predicates from the 13th to the 20th century. In the legends ‘STA’ stands for IL-states, ‘ACT’ for activities, ‘ACC’ for accomplishments and ‘ACH’ for achievements.

Figure 1. From Sánchez-Marco (2012)

(5) a. Martín fue condecorado
    Martin was condecorated
b. La Marta és educada
    The Martha is educated
c. La Marta està cansada
    The Martha is tired.
d. Tengo el brazo roto
    I-have the arm broken
The Peter has arrived

The rest of the talk will be devoted to provide the missing pieces to this account:

a) Show that there has been no reanalysis in the grammatical category of the participles: i.e. if one adheres to current categorical views of lexical categories or parts of speech, it can be shown that all participles are and they have always been deverbal adjectives. That is, there is no distinction between adjectival participles and verbal participles. There is only one class of words that are syntactically and semantically adjectives created out of verbal roots via derivational morphology.

b) Show that all BE + participle constructions in Latin as well as in the modern Romance and Germanic languages are instances of copular BE constructions involving adjectives derived from verbal roots. This includes the BE passive constructions as well as the BE perfect constructions.

c) Show that the eventive interpretations associated with both BE and HAVE perfect and with BE “passives” are consistent with an analysis of the participles as adjectives when we take into account the lexical semantics of the verbal roots, the interaction of the tense/aspect morphology in BE and HAVE with the lexical semantics of those verbal roots, the modification of the constructions by different types of adverbials and the two possible interpretations of some of the participles as resultant states and target states Kratzer(2000). In essence, what I will try to demonstrate is that even adhering to the most strict formal linguistics approaches, a constructional analysis that allows for all of the semantic interpretations traditionally associated with what were assumed to be different constructions is not only possible but more accurate and desirable. Finally, I will show that this analysis:

a) helps to explain the gradience phenomena displayed by auxiliary selection in the different Romance and Germanic languages.

b) makes it possible to relate mechanisms of syntactic change to the better studied mechanisms of sound change. Given the view of syntactic change as constructional change that will be defended in this talk, it becomes easier to explain why both sound changes and syntactic changes typically display the well-known S-shaped curve that is displayed by many changes related to social phenomena that don’t have anything to do with language. Adopting an agent based perspective on the study of language we can study the competition between different constructions to convey the same interpretations in a similar way we study other social phenomena involving competition between different alternatives. Network theory can thus provide useful models that can help us understand how linguistic changes spread through time and space. While these two perspectives are not frequently found in combination, I will make extensive use of constructs and argumentation from formal linguists to
lend additional support for a constructional, gradient and agent based view of syntactic/semantic change and of language in general.

References


According to Meillet (1912), who distinguishes two major mechanisms for the constitution of grammatical forms, viz. analogy and grammaticalization, only the latter can create new grammatical categories. The hypothesis we want to argue for in this paper is that analogy plays a major role in the development of new grammatical categories. Evidence will be provided by a diachronic and comparative analysis of two newly created grammatical paradigms in Romance which were inexistent in Latin, articles and auxiliaries.

1. Grammaticalization and analogy: two distinct mechanisms

In his pioneering 1912 paper, Antoine Meillet makes the following crucial distinction between analogy and grammaticalization:

Tandis que l’analogie peut renouveler le détail des formes, mais laisse le plus souvent intact le plan d’ensemble du système existant, la ‘grammaticalisation’ de certains mots crée des formes neuves, introduit des catégories qui n’avaient pas d’expression linguistique, transforme l’ensemble du système. (Meillet 1958 [1912] : 133)

Analogy and grammaticalization are thus considered to be two essentially distinct processes, since only grammaticalization can create new grammatical categories, i.e. categories previously unexpressed in the language. In recent work, although the role of analogy in grammaticalization has been widely acknowledged (Fisher 2012), some researchers (e.g. Lehmann 2004) still distinguish between ‘pure grammaticalization without analogy’ and grammaticalization steered by analogy. Examples of the former include the grammaticalization of the numeral ‘one’ into an indefinite article and that of the demonstrative into a definite article.

Latin did have neither articles nor auxiliaries. All Romance languages however
develop, at a different rate (Lamiroy & De Mulder 2012) and to a different extent, both grammatical categories. These two grammatical categories have a functional similarity: the main function of articles within the NP is to ensure the anchorage of the referent in the situational or textual context; similarly, the raison d’être of auxiliaries within the VP, is to anchor the verbal situation in the situational or textual context by specifying its tense, aspect and modality.

Zooming in on the emergence and historical development of these two categories will allow us to take a stand in the ongoing debate on the role of analogy in grammaticalization. Three Romance languages will be considered here: Italian, French and Spanish.

Romance languages offer a privileged area of investigation for diachronic and comparative linguistics and hence, for general linguistics. On the one hand, linguists have at their disposal a nearly uninterrupted documentation of two millennia, which is extensively accessible by means of electronic corpora. This allows us, on the one hand, to set up a fine-grained analysis of all or most of the different steps of linguistic change and on the other hand, to compare languages belonging to the same genealogical family but with contrasting typological tendencies. For these two reasons, Romance languages are an ideal testing ground for verifying general hypotheses about language and language change.

2. Analogy: a driving force for grammaticalization

Fisher (2007) has made a major contribution to highlighting the role of analogy in linguistic change. She argues that analogy acts both on the syntagmatic and the paradigmatic axis and affects the linguistic sign in its double-edged nature of form and function.

We argue here that nominal and verbal phrases evolve towards a similar abstract pattern on the syntagmatic axis: anchorage in the textual or situational context is increasingly expressed by grammatical elements at the left of the nominal and verbal head, by articles and auxiliaries respectively. On the paradigmatic axis, there is for both categories a tightening of the paradigm, i.e. reduction of number of members of the class, restructuring of the paradigm in terms of a limited number of binary features, adjustment of semantic features in order to fit in the paradigm.

In our view this striking parallelism is due to the pervasive action of analogy. We thus show that analogy is as powerful as to create new grammatical categories, challenging Meillet’s fundamental distinction between grammaticalization and analogy.
3. Asymmetry between NP and VP

Despite the action of analogy in both cases, NP and VP did not evolve in a strictly parallel way, i.e. the evolution occurs earlier and is more radical in the case of the NP than in the VP. For instance in French, zero marking disappeared completely and marking of the N by a formal determiner became obligatory. For the VP on the contrary, a mixed system survives in all Romance languages: TAM marking is partially expressed by suffixes on the verb and in part by free morphemes preceding the verb.

4. Different rates in genetically related languages

The correlation between the development of the articles and the auxiliaries is all the more salient when we compare different Romance languages. Grammaticalization of articles is in a more advanced stage in French than in Italian, which in turn is in a further stage than Spanish; a similar scale is observed for the auxiliaries (Lamiroy & De Mulder 2012).

4.1. Articles

French, Spanish and Italian develop a definite article from the distal Latin demonstrative ille and an indefinite article derived from the unity numeral unus, which is also attested in its plural form (Carlier & De Mulder 2011, Carlier 2013). In French, a third article is created, combining the spatial preposition de ‘from’ and the definite article, which entails the elimination of the plural form (Carlier 2007). As to Spanish, exploratory occurrences of the partitive are attested in the medieval language, totally in parallel with those that emerged in Old French. However, the partitive did not grammaticalize into a full-fledged article, while the plural form of the article derived from the unity numeral is conserved. As to Italian, the partitive did develop into an article, but it remains optional in Modern Italian and is perceived as a regional feature of the North (Carlier & Lamiroy 2014). Interestingly, the pattern with the determiner in initial position of the NP extends to demonstratives and possessives. The evolution with respect to the possessives is completed in French, but not in Italian and Spanish, where there is still a double system of prenominal possessive determiners and postnominal possessive adjectives (Van Peteghem 2012).

4.2. Auxiliaries

All three languages developed temporal, aspectual and modal auxiliaries by grammaticalization out of Latin full lexical verbs (Heine 1993, Lamiroy 1999),
e.g. Fr. aller and Sp. ir, which both function as an auxiliary of future tense, originated in the Lat. motion verbs ambulare ‘to walk around’ and ire ‘to go’ respectively. In Italian, andare whose etymology presumably also goes back to (a suppletive form of) the motion verb vadere ‘to go, to proceed’, viz. ambitare, is mainly a modal auxiliary with deontic value. However, of all three languages, French has the most grammaticalized auxiliary system, which is reflected on the syntagmatic axis by a reduction of the diversity of syntactic patterns and on the paradigmatic axis by a tightening of the paradigm, i.e. a reduction of the number of verbs which belong to the class (paradigmatization in Lehmann’s (1982) terms, compared to a larger class in Italian and an even larger one in Spanish, a restructuring of the paradigm in terms of binary parameters, and an adjustment of the semantic features of these verbs in order to fit in the paradigm.

References
Empirical observations across the world’s languages have shattered the decade-long assumption that grammatical case should be explained in terms of innate knowledge (Croft, 1991; Dryer, 1997; Haspelmath, 2007), whether this knowledge takes the form of abstract categories, thematic hierarchies or a shared conceptual space. It is therefore crucial to understand the processes that may give rise to new case systems if we want to find solid explanations for them.

This presentation focuses on the role of analogy in the emergence of a case system. Analogy is widely accepted among historical linguists as an important mechanism in language change, but so far it has typically been conceived as a trigger for small, local changes that leave the overall grammatical system intact (see e.g. Meillet, 1921). Through multi-agent experiments, I will demonstrate that analogy may fundamentally restructure the grammar of a language.

More specifically, I will present experiments in which autonomous artificial agents engage with each other in language games about real-world events (Steels, 2004; van Trijp 2010, 2012). In these experiments, agents are provided with an associative lexicon (but no grammar) for describing events to each other. In order to avoid cognitive effort in semantic interpretation, the agents can invent new case markers for indicating a particular participant role (e.g. the “pusher” of a “push”-event), or they can recruit existing markers through analogical reasoning on event structures, grounded in the agents’ sensorimotor experience.

The results show that a case strategy based on analogy has a distinct selective advantage for communication over event-specific marking: general case frames require a smaller inventory size, they propagate more easily in the population because their larger distribution comes with increased frequency, they facilitate the interpretation of novel forms, and so on. More importantly, as a side-effect of exploiting analogy in locally situated interactions, coherent case systems emerge on the population level that look from the outside as if they are manifestations of a universal semantic map. Instead of resorting to innate knowledge, similarities across languages can therefore be explained in terms of convergent evolution.
References


SEMANTIC COMPETITION AND (INTER)SUBJECTIFICATION IN THE SYSTEM OF THE DUTCH MODALS

JAN NUYTS

Department of Linguistics, University of Antwerp, Prinsstraat 13 R204
B-2000 Antwerp, Belgium

1. Topic

This paper presents the results of an investigation into the diachronic semantic evolution of the Dutch modal auxiliaries kunnen ‘can’, mogen ‘may’ and moeten ‘must’, focusing on the interaction in them between processes of (inter)subjectification (in the sense of Traugott, 1989, 1995, 2010, Traugott & Dasher, 2002) on the one hand, and a process of competition for semantic ground on the other hand. (See also Nuyts & Byloo, submitted.)

2. Data

The semantic evolution in the three modals is investigated by comparing their semantic profile in 4 stages in the language’s history, viz. Old Dutch (OD, before 1200), Early Middle Dutch (EMD, 1250-1300), Early New Dutch (END, 1550-1650), and Present Day Dutch (PDD, after 1950). The analysis is based on 200 instances per modal per period (but for OD the samples include all instances found in the few remaining texts, which never amounts to 200). But for PDD two separate sets of 200 instances are used: one, like for the earlier periods, exclusively written (henceforth PDDW), and one exclusively spoken (PDDS). Given the quite different nature of spoken and written language, these two sets are kept separated in the analyses. The samples are selected randomly from the full range of digitally available text sources for each of the periods, yet is subject to a balanced concern for representativity and reasonable spreading for each period, geographically and in terms of text types and authors, and for comparability between the periods (with PDDS as an exception of course).
3. Findings

The data are summarized in Tables 1-3. The meaning categories are ordered from top to bottom in terms of increasing (inter)subjectification (in the interpretation of it discussed in Nuyts, 2012, Nuyts & Byloo, submitted). (All interpretations below are supported by statistical testing, by means of Fisher Exact and the Spearman Rank Coefficient of Correlation.)

Table 1. Meaning development of kunnen.

<table>
<thead>
<tr>
<th>OD</th>
<th>EMD</th>
<th>END</th>
<th>PDDW</th>
<th>PDDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| n   | %   | n   | %    | n    | %
| 'know' | 2 | 12 | 7 | 4 | 1 | 1 |
| dynamic-inherent | 9 | 53 | 120 | 60 | 87 | 44 | 55 | 28 | 33 | 17 |
| dynamic-imposed | 6 | 35 | 70 | 35 | 85 | 43 | 71 | 36 | 104 | 52 |
| dynamic-situational | 3 | 2 | 22 | 11 | 33 | 17 | 25 | 13 |     |    |
| deontic |     | 4 | 2 | 24 | 12 | 21 | 11 |     |     |    |
| epistemic |     | 1 | 1 | 10 | 5 | 4 | 2 |     |     |    |
| directive |     | 1 | 7 | 4 | 13 | 7 |    |     |     |    |
| total | 17 | 200 | 200 | 200 | 200 | 200 |     |     |     |    |

Table 2. Meaning development of mogen.

<table>
<thead>
<tr>
<th>OD</th>
<th>EMD</th>
<th>END</th>
<th>PDDW</th>
<th>PDDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>dynamic-inherent</td>
<td>9</td>
<td>16</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>dynamic-imposed</td>
<td>27</td>
<td>47</td>
<td>45</td>
<td>23</td>
</tr>
<tr>
<td>dynamic-situational</td>
<td>12</td>
<td>21</td>
<td>59</td>
<td>30</td>
</tr>
<tr>
<td>deontic</td>
<td></td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>epistemic</td>
<td></td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>volitional</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>directive</td>
<td>6</td>
<td>11</td>
<td>64</td>
<td>32</td>
</tr>
<tr>
<td>concessive</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>conditional</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>57</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 3. Meaning development of moeten.

<table>
<thead>
<tr>
<th>OD</th>
<th>EMD</th>
<th>END</th>
<th>PDDW</th>
<th>PDDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>dynamic-inherent</td>
<td>3</td>
<td>60</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>dynamic-imposed</td>
<td>29</td>
<td>15</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>dynamic-situational</td>
<td>5</td>
<td>3</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>deontic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>evidential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>volitional</td>
<td>44</td>
<td>22</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>intentional</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>directive</td>
<td>2</td>
<td>40</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>conditional</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>5</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>
It turns out that, in their development from Old Dutch onwards, mogen and kunnen do, but moeten does not, show a clear pattern of evolution in terms of (inter)subjectification. (Detailed discussion and interpretation will be offered in the talk.) But the developments in the former two modals also show clear signs of an effect of the fact that historically they largely share the same set of meanings: both kunnen and mogen are ‘weak’ modals, but moeten is a ‘strong’ modal. This appears to trigger some kind of competition, whereby kunnen gradually acquires these meanings and mogen gradually loses them through time—cf. Table 4 (‘m%’ = share in each meaning of mogen, ‘k%’ = share in each meaning of kunnen).

Table 4: Evolution in share of mogen vs. kunnen per meaning category

<table>
<thead>
<tr>
<th></th>
<th>OD</th>
<th>EMD</th>
<th>END</th>
<th>PDDW</th>
<th>PDDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>m%</td>
<td>k%</td>
<td>n</td>
<td>m%</td>
<td>k%</td>
<td>n</td>
</tr>
<tr>
<td>dyn- inh</td>
<td>23</td>
<td>77</td>
<td>18</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>dyn- imp</td>
<td>57</td>
<td>43</td>
<td>33</td>
<td>39</td>
<td>61</td>
</tr>
<tr>
<td>dyn- sit</td>
<td>100</td>
<td>0</td>
<td>12</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>deo</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>3</td>
<td>56</td>
</tr>
<tr>
<td>epi</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>4</td>
<td>83</td>
</tr>
<tr>
<td>dir</td>
<td>100</td>
<td>0</td>
<td>6</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

There is no comparable competition for semantic ground in moeten.

4. Discussion

These observations strongly suggest an interaction between the element of semantic competition—which may be considered an effect of the principle of isomorphism/no synonymy—and the process of (inter)subjectification, whereby the former may actually be the trigger of the latter. In other words: although the (inter)subjectification hypothesis explains quite well the semantic developments in the modals when they occur (viz. in kunnen and mogen, but not in moeten), these processes are most probably sensitive to other diachronic forces, quite notably forces pertaining to the mutual effects of forms in a linguistic system, including, e.g., analogy (see Nuyts 2013 on the role of analogy in (de-)grammaticalization processes in the system of the Dutch modals), or, in the present case, the principle of isomorphism/no synonymy.

Even if we have no indications that these forces affect the actual course of the (inter)subjectification process, they quite likely do affect whether it happens or not.
Maybe the specific interaction of forces observed in the present case study should not come as a surprise, at least not if one adopts a functionalist perspective on language: the principle of isomorphism is very directly related to basic elements of communicative efficiency (avoid semantic unclarity), but (inter)subjectification is much less so, hence it would only seem ‘functionally logical’ if the former is more ‘agentive’ than the latter in shaping the linguistic system. This line of thought will be explored further in the actual presentation.

References
Nuyts, J., & Byloo, P. (subm.). Competing modals: Beyond (inter)subjectification. Diachronica.
ENGLISH *WH*-RELATIVES: TOWARDS AN AETIOLOGY OF A GRADUAL SYNTACTIC CHANGE

ROBERT TRUSWELL
Department of Linguistics, University of Ottawa, 70 Laurier Ave East
Ottawa, K1N 6N5, Canada
rtruswel@uottawa.ca

NIKOLAS GISBORNE
Linguistics and English Language, University of Edinburgh, Dugald Stewart Building
3 Charles St, Edinburgh, EH8 9AD, UK
n.gisborne@ed.ac.uk

We discuss a series of changes in the English relative clause system over c.1000–1500AD, and the prospects for modelling those changes. We show that subtle differences in the linguistic analysis can substantially affect the approach to modelling, and argue for a new understanding of these changes.

1. *Wh*-relatives and Indo-European

Headed *wh*-relatives, like (1), have a phrasal filler which reflects properties of the gap site. In contrast, fillers like *that* in (2) are monomorphemic and indeclinable. These contrasts motivate an analysis where *whose brother* in (1) moves from the gap site to [Spec,CP], while *that* in (2) is base-generated in C⁰.

(1) [NP The guy [CP *whose brother [IP I met _]]] was charming.
(2) [NP The guy [CP *that [IP I met _]]] was charming.

We call elements like *whose brother* relative specifiers. These are a largely Indo-European phenomenon: De Vries (2002) shows that 67.5% of IE languages have relative specifiers, but only 5.3% of non-IE languages. Although his sample omits several well-known cases of relative specifiers, for example in Finno-Ugric languages, they are still clearly concentrated in IE languages.

However, IE languages did not inherit their relative specifiers from Proto-Indo-European. Rather, the analogues of today’s headed relatives are adjoined
relatives, either clause-initial and marked with *kwɑ-*/*kwɑ*-i- or clause-final and marked with *yw- (Clackson, 2007). Neither type is embedded within NP, unlike (1)–(2).

In sum, relative specifiers are largely confined to IE, but not because of direct inheritance from PIE. Rather, other properties of IE languages make it particularly likely for learners of IE languages to introduce this construction, leading to parallel evolution of similar systems in several IE languages. By exploring this phenomenon, we hope to learn about the biases which predispose learners to introduce this construction into certain grammars, but not others.¹

2. English *wh*-relatives

English has had two sets of relative specifiers in its history. Until c.1200AD, English could form headed relatives using inflected demonstrative phrases (3); while the modern headed *wh*-relative system emerged slowly over c.1150–1500AD.

(3) Her feng to Dearne rice Osric [pone Paulinus ær
   Here succeeded to Deira kingdom Osric [that.ACC Paulinus earlier gekullode]
baptized
   “In this year Osric, whom Paulinus had earlier baptized, succeeded to the
   kingdom of Deira” (Peterborough Chronicle, 12th century, Allen, 1977)

Several differences exist between headed relatives with filled specifiers and without. Most importantly for us, relatives without filled specifiers contain only NP gaps, while relative specifiers can correspond to gaps of a range of categories (Allen, 1977).

The genesis of headed *wh*-relatives OldEnglish (OE) used *wh*-phrases as indefinites, and in questions and generalizing free relatives, inheriting all three functions from PIE *kwɑ*-/*kwɑ*-i-. English headed *wh*-relatives most likely developed out of postposed free relatives: there are several examples which are

¹ Comrie (1998) claims that relative specifiers are a European, rather than IE, phenomenon: they occur in Finno-Ugric but are rare in Indo-Aryan. This implicates constructional borrowing in their distribution. However, there are clear indications that relative specifiers have repeatedly evolved in parallel among genetically related languages. For example, English and French developed relative specifiers at roughly the same time, but neither borrowed the construction from the other.
both syntactically and semantically indeterminate between analysis as appositive generalizing free relatives and extraposed definite headed relatives.

(4) and eow ealle þing geswutelað, [swa hwæt swa ic eow seege] and you all thing show so what so I you say “and [he] explains everything to you that I tell you”, or “and [he] explains everything to you, whatever I tell you”

(Ælfric Homilies, late 10th century)

Semantically, this is a case of quantificational variability (Caponigro, 2003): it makes no difference to the interpretation of (4) whether hwæt is definite or universal. Syntactically, the ambiguity reflects a competition between two analyses: either the final relative in (4) is a (nominal) free relative in apposition to ealle þing, or it is an extraposed headed relative modifying ealle þing. The frequency of clause-final wh-relatives with nonadjacent antecedents (c.14% of wh-relatives) is intermediate between extraposed CPs (c.30% of adnominal CPs) and NPs in apposition (c.6% of adnominal NPs). This could suggest to a learner that not all wh-relatives are nominal; that some are headed relatives. This provides a plausible basis for the genesis of headed wh-relatives.

The spread of headed wh-relatives Headed wh-relatives initially had only oblique or adverbial gaps, complementing that-relatives with argumental NP gaps. Headed wh-relatives with argumental NP gaps initially occur in the 14th century, c.200 years after the first oblique headed wh-relatives (Fig. 1).

![Figure 1. Frequency of wh-relatives by grammatical function in Middle English](image)

3. Wh-relatives and Indo-European

The above description differs from the received wisdom in two ways. First, it downplays the similarity between different types of headed relatives: rather than asking “What can English speakers use to form headed relatives”, we ask “What
do English speakers do with \textit{wh}-phrases". Second, we interpret Fig. 1 as showing two discrete changes, while previous accounts (Romaine, 1982) have construed this as a single, gradual progression of a \textit{wh}-relative construction up Keenan and Comrie’s (1977) Accessibility Hierarchy. As the Accessibility Hierarchy is related to processing ease (Hawkins, 2004), the received wisdom therefore suggests a functionalist account, perhaps focusing on communicative need and processing ease.

However, functionalist accounts run into the problem that most of the world’s languages do not have relative specifiers, and do not need them. Any functionally motivated bias in favour of this construction must therefore be very weak, or headed relatives with filled specifiers would be typologically more common. It is also unclear why specifically \textit{wh}-phrases were co-opted for this purpose, and so often in IE. The present approach removes these obstacles, by demonstrating clear links with PIE *\textit{k\'o-}*/\textit{k\'i}-forms and OE \textit{hw}-forms.

Subtle refinements of the empirical picture therefore significantly affect our analytical and modelling options. On our account, the genesis of \textit{wh}-relatives reduces to an instance of choosing between two competing structural analyses of a surface phenomenon, a classic application of Bayesian reasoning. Meanwhile, the spread of \textit{wh}-relatives plausibly reflects the tension between various learning biases. For example, learners are biased towards associating a single form with a single function (\textit{mutual exclusivity}, Markman & Wachtel, 1988). This predicts that learners are biased against extending \textit{wh}-relatives to functions clearly associated with \textit{that}-relatives. The subsequent, possibly analogical, spread from oblique to argumental \textit{wh}-relatives exemplifies the general problem of how tightly a learner’s grammar should fit the input. Again, this is a classic Bayesian problem. An improved empirical description therefore leads to a more tractable modelling challenge.

References


1. Introduction

In this work, we model the historical development of verbal cluster order in Germanic languages. While there is an ongoing debate on the syntactic structure of these clusters, we created a simple model of surface patterns in which we view each order as a separate outcome, with a probability distribution over the outcomes. Using this model, we show that the current order in German and Frisian verbal clusters may have developed partly due to increased grammaticalization of embedding over time, and increased grammaticalization of tense and aspect. Subordinate clauses (to express embedding) and verbal clusters with ‘have’ (to express tense and aspect) became more frequent, leading to changes in verbal cluster word order.

2. Verbal clusters

The basic word order of the Proto-Germanic language is generally assumed to be Object-Verb (OV), though both OV and VO orders were probably possible. Modern Germanic languages adopted one variant without much variation — OV in the case of German, Dutch and Frisian, or VO for English and the Scandinavian languages. While all of these languages have verbal groups, in the OV languages they take the form of impenetrable verbal clusters at the end of clauses. For these clusters, different word orders are attested (Wurmbrand, 2006). We will limit our discussion to two-verb clusters, in which the finite verb can be positioned before or after the infinite:

(1) Ik denk dat ik het heb begrepen.
   I think that I it have understood
   ‘I think that I have understood it’

(2) Ich denke, dass ich es verstanden habe.
   I think that I it understood have
‘I think that I have understood it’

In the literature, construction 1 is called the ascending order (1-2 order or green order), and construction 2 is called the descending order (2-1 order or red order). In English, the Scandinavian languages, and sometimes in Dutch (as in 2), the ascending order is used (I have understood), while the default form of Frisian and German is the descending order as in 1. We limit our analysis to the OV languages, because the verbal groups in the VO languages are technically not clusters, and the descending order hasn’t been attested in any VO language.

In this work, we will model the diverging development of verbal clusters in these languages using an agent-based model, taking a reconstruction of the state of verbal clusters in Proto-Germanic as a starting point.

3. Model structure

We define a basic model of verbal clusters in terms of realizations with production probabilities. The model structure is based on the bidirectional model in Versloot (2008), though our models learn by interacting rather than iterating. Several instances of the model (agents) exchange verbal cluster realizations with each other, changing the probability distributions. The verbal cluster realizations depend on two features: construction type and construction context. Our model has three different construction types, reflecting the historical sources of verb clusters:

1. modal + infinitive: the origin of verb clusters in Germanic
2. ‘to have’ + PP: arose only later in history to extend the possibilities of expressing temporal and aspectual features
3. PP and copula + PP: originally a passive, predicative, construction — not purely verbal, rather adjectival.

As construction contexts, we consider main clauses and subordinate clauses, which differ in their word order in some Germanic languages. Furthermore, two realizations are possible for each of these constructions: the ascending and the descending order. Table 2 shows the structure of this model. Instances of the model produce exemplars of verbal clusters according to one of the two realizations.

We initialize the models with (relative) frequency figures that we reconstructed for 6th century Germanic, based on a comparison of Old English, Old High German and Old Frisian. Furthermore, Germanic languages have shown an increase over time of the number of subordinate clauses and the number of ‘to have’ + PP construction types. We simulate this by increasing the totals for these features (proportional over the two realizations). The model is run by having two model instances (A and B) exchange realizations. Model A produces a realization of a construction according to its probability distribution, and subtracts it from its frequency figures (it is given away). Model B then adds to its stored frequencies,
and replies with a realization according to its own probability distribution. The models thus develop their probability distributions in the same way. This happens even though the models are only exchanging realizations, no information on the probabilities of individual features encoded in that realization.

The models converge from their predefined, proto-Germanic probability distribution to a state in which probabilities are distributed based on the features of the model. We then compare the resulting model outputs with actual Germanic language texts to see how well we have modeled the real state of these languages.

4. Realization probabilities

A model’s realization probability is based on the probabilities of its features. An ascending realization may be produced according to the following:

\[ P(\text{asc}|x) = P(\text{asc}|x_{mc}) \times P(\text{modinf}|x_{modinf}) \]

where \( x \) is a set of feature values. \( P(\text{asc}|x_{mc}) \) represents the level of ambiguity of the ascending order as a main clause, \( P(\text{asc}|x_{modinf}) \) for the modal+infinitive construction type. These probabilities are calculated from the stored frequency of the features in ascending contexts, i.e.

\[ P(\text{asc}|x_{mc}) = \frac{F(\text{mc},\text{asc})}{F(\text{mc})} \]

The results from mixed combinations (\( P(\text{asc}|x_{mc}) \times P(\text{desc}|x_{sub}) \), etc) are neglected. The effect is that constructions which are relatively unambiguous because they show a strong form-function correlation, are favoured, others are disfavoured, i.e. the different features that these realizations encode, become well represented in the two forms.

5. Results

| Table 1. Descending output probabilities from early–Modern Frisian text c. 1550. |
|---------------------------------|----------|----------|----------|
| %descending | mod+inf | habba+PP | cop+PP   |
| main         | 100%    | 100%     | 100%     |
| sub          | 100%    | 33%      | 20%      |

| Table 2. Descending output probabilities of model. |
|---------------------------------|----------|----------|----------|
| %descending | mod+inf | habba+PP | cop+PP   |
| main         | 100%    | 92%      | 70%      |
| sub          | 98%     | 33%      | 9%       |

We compared the model to frequency figures for early-Modern Frisian (ca. 1550) once the proportions main clause–subordinate clause and the proportions
between the three constructions were comparable to those in our Frisian dataset (table 1 and 2). As such, the results are promising. Tuning of the model – e.g. by a slower or quicker rise of the amount of subordinate clauses — shows that it in the long run it tends to produce 100% ascending or 100% descending realizations for all feature sets. The situation with 100% descending realizations reflects basic word order in German and Frisian, although then V2-movement is needed to get the finite verb in the second position in main clauses. We assume V2 to be a grammaticalised side effect of the asymmetries as reconstructed for Proto-Germanic, where ascending orders were dominant in the combination of modal+inf, which happened to occur more often in what we call main clauses from the modern perspective than in subordinate clauses. The current model is probably too crude to model more complex word orders.

According to our model, different speed of grammaticalization of have+pp and increase of subordinate clauses (both represented by increased frequency) may affect the balance between ascending and descending orders. The descending order is supported by the grammaticalization of embedding. Due to V2 movement in these languages, the finite verb precedes the other verb in main clauses. This ascending order differentiates main clauses from subordinate clauses, motivating the preservation of a descending order in the subordinate clauses. Increased use of subordinate clauses may then have supported the descending order as the base order. However, if have+pp grammaticalizes earlier, the ascending order is supported. Other syntactical or stylistical differences between languages may also explain whether a language moves towards ascending or descending orders.

Our model cannot yet account for the current state of the Dutch language, which first moved towards mainly descending orders like German, and then shifted towards ascending orders again, a change that is still in progress, considering the current state of variation (example 2 and 1) (Coussé, 2008). There is evidence that the ascending order has become the default form (Evers, 1975), and this second change was likely caused by a factor outside the scope of our model.

Overall, it can be concluded that the interaction of basic probabilistic choices of constructions with shifting input and shifting preference of constructions may be a key to understanding different word orders in the Germanic languages.

References

Language is a system of behaviour that is shared by members of a speech community. A key question is how the collective dynamics of language (e.g., changes in grammatical structure) are shaped by individuals’ cognitive apparatus and interactions between speakers. Can certain linguistic structures and changes be assigned primarily to universal factors, or are culturally-specific factors also at play? I will discuss how mathematical models may be used to help answer such questions. I will focus on the intuition gained from modelling complex systems in the physical sciences in identifying the key drivers of a collective phenomenon, how the resulting models can be related to theories in linguistics, and how to make effective use of the sparse data that is typically available for historical language change processes.
WHEN DO CREOLE LANGUAGES EMERGE?

FRANCESCA TRIA

Institute for Scientific Interchange (ISI), Via Alassio 11C, Torino, Italy, 10126

In the framework of language games I will investigate the emergence of a new language out of the contact of two preexisting ones. In particular, I will show how a simple variant of the Naming Game, enriched by a suitable contact ecology, can predict in what conditions there is the emergence of creole languages in surprisingly agreement with real data.
We present a longitudinal corpus-based analysis of 15 authors writing in the 17th and 18th century in an ongoing grammaticalization process (c.q. the auxiliarization of be going to). Our aim is to arrive at a fine-grained analysis of the micro-changes involved in this kind of language change. In this way, we hope to help bridge the gap between the agent-based modelling and the more traditional grammaticalization studies.

1. Intro: Breaking down the aggregate view on grammaticalization

Investigating individual differences in language behaviour by looking at non-elicited ‘naturalistic’ data has recently been made easier by the increasing availability of large-scale corpora, especially for English (Barlow 2013). Recently, some interesting diachronic corpus studies in this field have been carried out (Nevalainen et al. 2011; De Smet, ms.), to arrive at the constraints individual variation is subjected to. These studies, however, do not take a longitudinal perspective, in which individuals are followed through time, to see how they shift their behavior, accommodating to or diverging from particular ongoing changes. The few longitudinal studies that we have (Bergs 2005; Raumolin-Brunberg 2009; Hendriks 2013), are typically small-scale. The present study tries to combine the longitudinal approach with large-scale corpus analysis. We present longitudinal individual data on what is perhaps the most iconic of grammaticalization cases: the rise of be going to as a marker for future in English. We make use of the large-scale EEBO corpus, to see how individual
languages users behaved in the seventeenth and eighteenth centuries, a crucial period in the evolution of going to.

Breaking down the aggregate view on the grammaticalization of be going to into individual users' behavior may help bridge the gap between ‘traditional’ diachronic linguistics and agent-based modeling. Agent-based models (Steels 2011) are able to show how emergent properties of language structure arise from well-defined individual interactions (Landsbergen et al. 2010; Beuls & Steels 2013), but are sometimes criticized for the allegedly artificial nature of the communicative setting. On the other hand, traditional corpus-based diachronic linguistics often fail to specify the precise conditions of naturalistic settings between real-life agents partaking in ongoing language changes. At present, it is debated whether adults, adolescents or children are the main instigators of language change. Some scholars argue that language change primarily happens over generations (e.g. Lightfoot 1999), while others argue that it takes place during lifetime (Croft 2000; Bergs 2005).

2. Methodology

2.1. Corpus description and data extraction

In order to examine if micro-steps in the grammaticalization (or grammatical constructionalization Traugott & Trousdale 2013) of be going to occur within real-life agents’ lifetimes, we selected 15 prolific authors from EEBOCorp 1.0 (Petré 2013), a half billion+ corpus based on the EEBO-database (eebo.chadwyck.com), containing English books printed between 1473-1700. Selection criteria were: (i) Sufficient material is available for the first and second halves of writer’s careers; (ii) Constant register over time; (iii) Writers are from roughly the same social status. Posthumously published works in EEBO not included in EEBOCorp 1.0, and translations done by one of the selected authors were also included. The post-1700 output of Burnet, D’Urfey and Dunton was added from the Eighteenth Century Collections Online database (ECCO).

The resulting corpus consists of about 31 million words, with individual author word counts ranging between ca. 300,000 and 10,000,000 words. All forms of going were extracted from this corpus by means of Perl scripts (n = 5821), taking into account variant forms identified in an exhaustive token list.
Additional scripts and manual analysis was used to filter out a total of 1591 instances of *be going*.

### 2.2. Data coding and analysis

We coded the EEBO datapoints for several formal and semantic features that are commonly associated with the grammaticalization of *be going to*, and can serve as diagnostics to assess the level of grammaticalization reached in a particular individuals, which serves as the dependent variable in our inquiry. Each of these features is analyzed with a level of granularity that allows us to pick up small increments in the level of grammaticalization. In the analysis, we both looked at the behaviour of each feature separately, and at their combined value, by computing a summative measure of grammaticalization. For each of the authors, we divided the collected data in half, to arrive at two categories ‘earlier work’ and ‘later work’, in order to check whether differences occurred through the years.

### 3. Findings

The scatterplot in the left panel in Fig. 1 brings out the aggregate view on grammaticalization: the score on the Y-axis is a summative measure of how many grammaticalization features a certain datapoint displays. The regression line (lowess) has an s-shaped curve, typical of language change. The rise is significant (Kendall tau = 0.126, p < 0.0001 – the relatively weak effect size is not surprising, considering that we only look at a time window of 50 years). The right panel breaks the data down into the two periods for each author. Authors with an increased grammaticalization score in their later work are indicated in red. As can be appreciated, they form the majority of the individuals investigated. Overall, we see an increase in grammaticalization scores through time (lowess regression line). In our paper, we will investigate the differences between the authors in depth.

### References


Figure 1: Grammaticalization of *be going to* in EEBO
Posters
There is a wide consensus in evolutionary linguistics that the evolution of language is constrained by the cognitive capabilities of language users, who must be able to acquire and process language given the limited resources they have at their disposal. It is therefore crucial to analyze the computational complexity of possible language systems in order to explain why we might (or might not) expect such systems to emerge in a speech community.

This poster presents such complexity analyses for language comprehension, using an agent-based model of cultural language evolution in which a population of autonomous artificial agents engage in multireferential language games with each other (Steels and Casademont, 2013). In our experiments, we first show how these agents can self-organize four different language systems: a lexical (pidgin) language, a word grouping language, a sequencing language, and a patterning language (which all four progressively scale towards the kinds of constituent structures found in most human languages). Through a complexity analysis of semantic interpretation, we show that each system progressively increases interpretation efficiency, which may explain why almost all human languages have evolved constituent structure.

References
Most languages either have case marking, adpositions, or both. Languages with case marking can 'lose' their case system, which is then generally replaced by adpositions. Despite the enormous amount of work done on the subject, we still cannot figure out completely how the process of case-loss takes place (Hagège, 2010). In this research we present two agent-based models of case-to-adposition change in languages, inspired on our experience (Fagard, 2010) in diachronic studies of cases and adpositions in languages. The first model is a simple approximation that uses neural networks to model the introduction of adpositions to desambiguate ambiguous sentences. Based on the results obtained and the limits of our first model, we present the design of a new experiment inspired on cultural language evolution experiments (Steels, 2012; van Trijp, 2010) to model the process of grammatical change in case marked languages.

References

MODELLING THE ROLE OF MOTION VERBS IN THE EVOLUTION OF RUSSIAN ASPECT

YANA KNIGHT
Vrije Universiteit Brussel, Artificial Intelligence Laboratory
yana@ai.vub.ac.be

MICHAEL SPRANGER
Sony Computer Science Laboratory, Tokio
michael.spranger@gmail.com

Motion verbs occupy a special place in Russian. It is believed that historically such verbs grounded in embodied movements served as prototypes for other, non-motion verbs by providing a concrete ground as events that allowed profiling of the beginning, middle and end of the motion, as well as its goal-directedness (Janda, 2008), (Janda, 2010). In non-motion verbs goal-directedness was extended to completability - whether a verb could be associated with a result - and became expressed with a pair of distinct verb stems describing the same event that differed only in aspect (Janda, 2008).

While the majority of the verbs eventually lost such lexical stem differences through grammaticalisation of aspect, lexical aspect was retained in around a dozen motion verbs in Modern Russian (Janda, 2008). It is believed that these verbs of motion, through their prototypical role, facilitated the transfer not only of goal-directedness from motion to non-motion verbs but also of lexical aspect to grammatical thereby reducing the number of verbs with distinct stems (Janda, 2008). Possibly due to their importance as prototypes, they preserved their own lexical aspect marking.

The present work is part of a larger project dedicated to modelling this historical phenomenon. We intend to accomplish our goal by initially re-constructing the current state of verbs of motion in Russian and in the later stages, tracing historical developments through modifications to the current grammar. At the present stage, we demonstrate a grammar of Modern Russian verbs of motion implemented with the Fluid Construction Grammar formalism (Steels, 2011a). In future experiments, agent-based models and language game paradigm (Steels, 2011b) will be employed in order to simulate the grammar and its historical development in use.
References


ENTRENCHMENT VS. TRANSPARENCY. MODELLING THE DUTCH STRONG-WEAK PAST TENSE COMPETITION IN AN AGENT-BASED SIMULATION.

DIRK PIJPOPS, KATRIEN BEULS
Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, Brussels, B-1050, Belgium

FREKK VAN DE VELDE
QLVL, Katholieke Universiteit Leuven, Blijde-Inkomststraat 21, Leuven, B-3000, Belgium

Typically, Germanic verb inflection falls apart into two main classes: strong and weak verb inflections. At present, the strong verb inflection system has lost most of its original transparencies and disintegrated into an intricate patchwork of distinct classes and irregularities (Haeseryn et al., 1997, pp. 87-94; Lieberman et al., 2007; Mailhammer, 2007). Yet, although the alternative “weak” inflection system – which also developed quite early - is much more transparent and highly productive, the strong verb inflection continues to show itself remarkably resilient. Not only has it resisted the strong regularization pressure of the weak verbs relatively well, it incidentally even shows some signs of expansion (Salverda, 2006, pp. 170-179).

In order to investigate how such an untransparent system can survive and even incidentally expand in a population, we have constructed an agent-based model of the competition between the strong and weak verb forms in Dutch. In our current model, the agents are embedded in a world of events, which they need to communicate to one another in a language game (Steels, 1995). These events are typically expressed by strong verbs in Dutch and their frequency correlates with the frequency of the verbs describing them in the Corpus of Spoken Dutch (CGN, cf. Van Eerten, 2007). The more often an agent
hears the strong or weak form of a particular verb, the more likely he is to use this form in a future game. While the agents start with an outspoken preference for the strong forms – corresponding to the current situation in Dutch – the weak forms benefit from being more transparent. That is, while the use of a strong form only affects its direct counterpart in the lexicon of the hearer, the use of a weak form also slightly raises the probability of all other weak forms in the hearer’s lexicon due to the transparency of weak inflectional endings.

Although the current state of our model is too simple to accurately model the historical competition, it is our aim to ultimately compose a truly realistic model. To achieve this, we mean to go as far as possible in incorporating the vast body of knowledge already available on the strong-weak verb competition.

References
EMERGENCE AND (CO)-EVOLUTION OF TENSE, ASPECT AND MODALITY

PAUL VAN EECKE
Sony Computer Science Laboratory Paris, 6 rue Amyot, Paris, 75005, France
vaneecke@csl.sony.fr

Language users of different languages employ many different strategies to express tense, aspect and modality (Comrie 1976, 1985). This poster presents the first results of a research project that aims to show how a population of language users may self-organize such tense-aspect-modality (TAM) systems from scratch through agent-based modeling.

More concretely, we will present a computational reconstruction of the Dutch TAM-system in Fluid Construction Grammar (Steels 2011a, 2012) that works for both parsing and production. The reconstruction demonstrates that TAM-systems of human languages go well beyond simple associative communication systems in which there is a one-to-one mapping between meaning and form. Instead, grammatical TAM-systems consist of an abstract and hidden layer of semantic and syntactic categories that mediate between rich conceptualizations and their morphosyntactic realization.

In future work, we will incorporate our processing model in agent-based experiments based on the language game paradigm (Steels 2011b). This work will proceed in a stepwise fashion, whereby first the necessary learning mechanisms are operationalized that enable autonomous artificial agents to acquire a sophisticated real-world TAM-system. Secondly, we will investigate how agents can self-organize their own TAM-system of human language-like complexity.

References