

Changing Patterns of Deixis in the Nominal Domain¹²

Alexandra Simonenko

Ghent University, 9000, Ghent, Blandijnberg 2, Belgium

*alexandra.simonenko@ugent.be

Abstract:

This chapter discusses theoretical issues related to historical developments of adnominal demonstratives against the backdrop of a methodological paradigm transition from the so-called etymological argumentation to arguments based on the observation of changing quantitative patterns in the historical data. The methodological shift challenges some of the previously held empirical assumptions and calls for new theoretical answers. Using better documented languages for case studies, I show that recent quantitative results challenge the long-held view about historically “observable” cyclic shifts from demonstratives proper to definite articles. In particular, the results points to a greater diachronic stability of demonstrative systems than was assumed based on the etymological arguments.

Keywords: nominal deixis, demonstratives, strong definite articles, Greenberg’s definite article cycle, semantic change, diachronic stability

0.1. Introduction

The term *deixis* (from Greek *deiknumi* “I point out”), already used by ancient Greek grammarians, reappeared in Brugmann 1904 in the context of the discussion of the properties of demonstratives in Germanic languages. Bühler (1934) further develops this notion, incorporating it into his theory of communication, whereby he distinguishes between two complementary but phenomenologically and psychologically different aspects of language: a deictic field and a symbolizing (naming) field, the former consisting of context-dependent elements such as pronouns like *I* and *you*, demonstratives like *this* and *that*, and temporal expressions like *now* and *then*, whereas the latter includes for instance nouns and verbs, which refer to objects and events in a way that does not change depending on the context. Bühler (1934) is the first one to come up with the

now commonly used metaphor of a system of coordinates to capture the role of deictic expression in the propositional meaning. According to Bühler (1934), the meaning of the words *I*, *here*, and *now* determines the center, *origo*, of a person-space-time coordinate system relative to which the meanings of other deictic terms are established. Bühler (1934) distinguishes between three psychological planes within which a coordinate system can be established: perceptual reality, imagination, and discourse. The three corresponding types of deixis or modi of pointing (*Zeigemodi*) are visual demonstration (*demonstratio ad oculos*), imagination-oriented deixis (*deixis am phantasma*), and anaphoric deixis. In the visual demonstration mode the *origo* is perceptually accessible, while in the imagination-oriented mode the *origo* corresponds to the person, time, and location of an imaginary protagonist. This mode also subsumes cases whereby the person component of *origo* corresponds to the speaker (rather than an imaginary protagonist), while the space and time are imaginary, rather than perceptually accessible dimensions. The replacement of the speaker by an imaginary protagonist as the person coordinate of *origo* is picked up in (Lyons, 1977, 579) under the term *deictic projection*. Finally, in the anaphoric mode, the *origo* corresponds to a moving point in the ongoing discourse. Similarly, Fillmore (1971) distinguishes three ways of *anchoring* (a term initially introduced by Rommetveit (1968)) the meaning of a deictic word: gestural, symbolic, and anaphoric. Fillmore (1971, 44) notes that deictic expressions which are limited to gestural anchoring appear to be rare, mentioning that French *voici* and *voilà*, as well as English *yea* (*She's about yea tall*), which he calls “presentatives”, seem to fit this profile. Already Bühler (1934), and later Lyons (1977) assume that most deictic elements involve a combination of symbolic and deictic components, with the exception of particles. Based on

the absence/presence of a symbolic admixture, Lyons (1977) distinguishes between *pure* and *impure* deixis, respectively. Diessel (1999) subsumes anaphoric uses under the term endophoric, together with discourse deictic (propositional anaphora) and recognitional uses, which are contrasted with exophoric uses involving reference to entities in the discourse situation.

The three anchoring mechanisms can be illustrated with different uses of English *there*. (1) illustrates gestural anchoring, since the hearer “ha[s] to know where the speaker is pointing in order to know what place he is indicating”, Fillmore (1971, 41).

- (1) I want to put it there. [gestural anchoring]

Fillmore (1971, 41) gives (2) as an example of symbolic anchoring, whereby “we understand the word ‘there’ as meaning “the place where you are”. In Bühler’s terms, while the person component of the *origo* corresponds to the actual speaker, the spatial dimension of the coordinate system, crucial for the interpretation of *there*, is not perceptually accessible and lies entirely in the domain of mental representations.

- (2) Is Jonny there? [symbolic anchoring]

Finally, (3) is a case of anaphoric anchoring, with *there* referring “to a place which had been identified earlier in the discourse”.

- (3) I drove the car to the parking lot and left it there. [anaphoric anchoring]

Fillmore (1971, 38) explicitly states that an utterance containing deictic ele-

ments can only be interpreted correctly if these deictic elements are anchored onto appropriate parameters of the utterance context, such as the participants in the communication act, “location in which speech act participants are situated”, “time of message encoding and message decoding”, discourse (“linguistic material within which the utterance has a role”), and social aspects (“the social relationships on the part of the participants in the conversation”). Correspondingly, “aspects of language which require this sort of contextualization” are called deictic. On this view, deictic elements are part of a larger family of expressions whose interpretation interacts with the environment in which they are uttered.

Although the common semantic denominator of deictic expressions can be defined as the property of having their interpretation linked to certain parameters of the extra- or intra-linguistic context, deictics used to refer to the speech act participants seem to stand apart from the rest with respect to how the referential connection comes about. While Bühler (1934, 109) notes that in the use of *I*, the quality of the speaker’s voice can serve as a “pointer” to the identity of the speaker, akin to a pointing gesture involved in the *ad oculos* uses of demonstratives (as in *Who is there? It is me*), it is also true that such uses are rare, since “speaker and hearer are usually aware of their communicative roles” (Diessel (2012b, 8)), which makes an identification procedure unnecessary. Moreover, it does not seem that identification in the course of the speech act is required for the grammatical use of such deictics. As Bühler (1934, 94) himself writes, “what characterizes the person identified by the word *I* is primarily nothing other than the role of the sender in the ... present exchange of signals”. That is, as the term *person* (from the Greek *prosopon* “mask, role”) suggests, first and second person pronouns do not require that the hearer were

able to perceptually or mentally (as in the memory-based deixis) identify the referent *in the course of the speech act*. The sender of the message “When you leave, please close the door” is perfectly happy knowing that the door will be closed by whoever reads the message, no matter who the person is. This is not the case for non-participant deictics, such as third person pronouns or demonstratives, which Benveniste (1956) called “non-person” pronouns. The use of such forms requires a recoverable connection with an extra-linguistic or a linguistic entity outside of the noun phrase in which it is used. Buridant (2000, 123) drawing on Kleiber (1990), assumes a distinction between transparent, complete deictics on the one hand and opaque, incomplete deictics on the other. The former, such as French *je* “I”, *tu* “you”, *aujourd’hui* “today” are self-sufficient for their interpretation, while the latter, to which the demonstratives belong, “cannot receive their interpretation based solely on their utterance” (translation is mine – author), but “can be linked to a gestural, anaphoric or symbolic use”. Wackernagel (1918) states that demonstratives “always contain a ‘Look over here!’” meaning component, while Diessel (2012b) argues that the use of demonstratives crucially involves attention manipulations. The referential connection between an anchor (a location) and the denotation of the noun phrase (an individual or object) occurs as a result of the speaker’s *directing* the attention of the hearer to the right spatial area. Demonstratives thus can be considered as inherently interactional deictics. It has been observed cross-linguistically that demonstratives tend to be topic-shifters and to resist referring to already established topics (Weinrich (1993), Zulaica-Hernández and Gutiérrez-Rexach (2011), Schwarz (2015), Fuchs and Schumacher (2020), Patil et al. (2020)). These observations go along with Wackernagel’s intuition about demonstratives as attention-callers, a role opposite to topic continuation.

This chapter is dedicated to incomplete, to use Kleiber’s term, deictic expressions occurring adnominally, that is, demonstratives and, to a lesser extent, adnominal locative particles. While person-anchored deixis is central in pronominal systems, as highlighted in WBCDL112, this type of deixis will only be discussed here as a possible outcome of the evolution of other types of adnominal deictic expressions. For the discussion of social deixis the reader is invited to see WBCDL138, and for time deixis WBCDL123.

Semantic changes can and often are accompanied by morphophonological ones, such as a change from a free to a clitic morpheme status etc. This chapter, focuses, however, solely on the semantic dimension. The term *deixis* is currently in competition with the originally Charles Pierce’s semiotic theory term *indexicality* (from Latin *index* “pointing finger”). In the literature that uses “indexicality” as the primary term, *deictic* (as in “deictic features of demonstratives”) is sometimes reserved as a shortcut for “spatially deictic”. This chapter aligns with Bühler’s terminological tradition.

In what follows, I will consider changes affecting adnominal demonstratives first from the perspective of shifts in the meaning (and distribution) of individual items, and then from the perspective of the determiner paradigm as a whole, illustrating both with case studies.

0.2. Evolutionary pathways

Deictic morphemes belong to the set of most universally attested grammatical elements, which is not surprising given that, in the words of Lyons (1995, 305), “most utterances ... in all languages are ... deictic, in that the true-value of the propositions they express is determined by the spatio-temporal

dimensions of the deictic context” or, in a much earlier formulation of K. Brugmann, “[e]veryday human contact is such that what the speaker says is understood by the addressee largely on the basis of the situation in which the utterance occurs” (cited from Bühler (1934, 98)). Lyons (1995) argues that the *deictic context*, defined as a system of coordinates with the zero coordinates determined by the identity of the speaker, the time and the place of utterance, is an integral part of the utterance context. Understanding the crucial role of context and context-dependent elements in communication and the central role of gestures and gesture-like signals in the use of deictics lead K. Brugmann to postulate that gestural communication is primordial in relation to verbal and, consequently, to assume deictic origins of natural language.

The view that came to dominate at the end of the twentieth and beginning of the twenty first centuries identifies the source of the grammatical elements in the lexical (i.e. Bühler’s symbolic) material (Bybee 2003, 161, Heine and Kuteva 2007). Heine and Kuteva (2007), in particular, mention motion verbs as possible source of demonstratives. This view is criticized in Diessel (2012a, 36) who argues that the phonetic similarity between demonstratives and the ‘go’ verb in a sample of three languages is not a convincing argument in favour of the lexicalist view. Diessel (2012a) emphasizes that, apart from the cases mentioned by Heine and Kuteva (2007) and related studies, there is no evidence that demonstratives, in particular, were at some point derived from lexical words. More generally, Diessel (2012a) revives the deictic hypothesis, bringing to the fore evidence that demonstratives are basic and serve as grammaticalization sources for a wide array of functional elements. Diessel (1999) identifies 17 “grammaticalization channels” that have demonstratives as their source, of which adnominal demonstratives account for 5

(definite articles, boundary markers of postnominal relative clauses/attributes, determinatives, number markers, and specific indefinite articles), pronominal demonstratives for 7 (third person pronouns, relative pronouns, complementizers, sentence connectives, possessives, verbal number markers, pronominal determinatives), adverbial demonstratives for 2 (temporal adverbs, directional/locational preverbs) and identificational demonstratives for 3 (non-verbal copulas, focus markers, and expletives). The 5 grammaticalization pathways of the adnominal demonstratives, which are the focus of this chapter, are briefly discussed below, with the most robustly attested pathway, “demonstrative > definite”, closing the list.

“**Boundary markers of postnominal relative clauses**”. Diessel (1999, 26) gives (4) from Sango (a creole spoken in the Central African Republic) as an example of pathway 2, where a demonstrative form *só* (the second instance) is used as a boundary marker of a relative clause, on the analysis of Samarin (1967, 73).

- (4) áfamille só ahé mbi só, mbi yí ála pepe
 relatives this laugh 1SG this 1SG like them NEG
 “Relatives who make fun of me, I don’t like them.” Samarin (1967, 73), cited from Diessel (1999, 26)

Sankoff and Brown (1976, 657) assume a similar pre-theoretical status for a Tok Pisin particle *ia*, etymologically originating in English *here*. The label is also used in the grammar of Bukawa (Austronesian) by Eckermann (2007) for the morpheme *naŋ* on its appearance following a relative clause, as in (5), where the relevant *naŋ* is glossed as *giv* (“relative clause boundary marker”) (RC stands for a “complementizer for relative clauses”, POT stands for a “potential modal adverb”, NR for a “non-realis modal adverb”, and IR for “irrealis”).

- (5) Asa naŋ bu nem gwelɛŋ naŋ oc tap ŋaôli sa.
 who RC POT 3S.IR.do work GIV NR 3S.find payment up/out
 “Whoever does work, he will get pay.” Eckermann (2007, 32)

“Determinatives”. Diessel (1999) suggests putting morphemes that look like definite articles used with noun phrases modified by a restrictive relative clause into a special category of determinatives. The rationale for this categorization is morphosyntactic idiosyncrasies associated with such a configuration, such as the omission of an otherwise required definite suffix in Swedish. More generally, it seems to be not unusual for a restrictive relative clause to have an effect on the patterns of definiteness marking. In particular, Germanic varieties distinguishing between the so-called weak and strong definite articles use the latter in the presence of a relative clause (e.g. Wiltschko 2012). It is however less obvious that this is a result of a separate diachronic development leading to the emergence of a special category rather than a consequence of the semantics of anaphoric demonstratives/articles themselves. In particular, Simonenko (2014) proposes that anaphoric demonstratives/articles, in virtue of their internal structure and semantics, relate the denotation of the noun phrase either to a previously mentioned individual or to the property denoted by a relative clause.

“Number markers”. There is some evidence that demonstratives may give rise to plural markers (WBCDL 133). Since the evidence rests on etymological, rather than “real-time” observations, it is unclear whether the development is direct or necessarily goes through the phase of third person plural pronouns, which, in turn, develop into markers of plurality.

“Specific indefinite articles.” Diessel (1999) sets apart uses of demonstratives, as in (6), which appear to pattern with indefinites, including not

giving rise to the definiteness effect in existential constructions (in the sense of Milsark (1974)). As in the case of “determinatives”, it is unclear whether this phenomenon calls for a special diachronic path, rather than being an effect of the general demonstrative semantics which can be considered under the umbrella of the so-called affective uses (Lakoff 1974), widely attested both diachronically and typologically.

(6) There was this boy in my school.

“Definite articles”. Of the 5 pathways identified by Diessel for adnominal demonstratives, only the pathway “demonstrative > definite” is uncontroversial (for seminal works from the twentieth centuries, see Diessel (1999, 25)), while the others are much less robustly attested and, from a theoretical standpoint, correspond to descriptive generalizations which, upon closer inspection, may eventually be subsumed under the definite article pathway. The “demonstrative > definite” shift, with which the case studies in this chapter are largely concerned, is taken by Greenberg (1978) as the first step in the diachronic sequence in Table 1.

Table 1: “The cycle of the definite article”, Greenberg (1978, 61)

Stage 0	Stage I	Stage II	Stage III
demonstrative	definite article	non-generic article	noun marker

In relation to the transition from Stage 0 to Stage I, Greenberg (1978, 61) hypothesizes an anaphoric stage, arguing that a definite article “develops from a purely deictic element which has come to identify an element as previously mentioned in the discourse”, an anaphoric use often being “an additional function of an element which is also a pure deictic”.³ He further notes that

“sometimes there is a particular demonstrative which has assumed [anaphoric use] as its basic function”, such a demonstrative most often being “one which points to a location near the third persons rather than the first or second person”. This hypothesis is reiterated in Hawkins (2004, 84) and a more recent study of the emergence of Mainland Scandinavian definite articles by Skrzypek (2012, 47), who assumes that “[t]he development from Stage 0 to Stage I can be further subdivided into sub-stages. It originates with the use of the demonstrative in anaphoric contexts, when an exophoric (situational) marker is used intra-linguistically (e.g. Lyons 1975, Diessel 1999, 109-111).” Similarly, Kuteva et al. (2019, 137) mentions that “[t]he [DEMONSTRATIVE > DEFINITE] pathway appears to be restricted to demonstrative forms having an anaphoric function.”

These assumptions are part of a more general consensus dating to the first works on the topic that exophoric deixis diachronically precedes other types of deixis. Bühler (1934) follows Paul (1880) in assuming that the visual demonstration precedes the development of anaphoric and memory-oriented deixis. Discussing Brugmann’s and Bühler’s contributions, Diessel (2012a, 45) concludes that “[t]he development of grammatical markers from demonstratives is based on the anaphoric use of spatial deictics whereby a genuine pointing word is utilized for a language-internal, grammatical function”, identifying the anaphoric mode of deixis “as a particular mechanism of change that shares important properties with the metaphorical use of content words which (often) underlies the incipient stages of the grammaticalization of symbolic terms”.

Adnominal morphemes with the basic function of establishing anaphoric reference have come to be known in the literature as strong articles since Ebert (1970). Comparing the patterns in languages for which such morphemes have been reported (Schwarz 2019 for a typological overview) and in those for which

no such claims have been made, it is clear that a theoretical distinction needs to be drawn between morphemes which allow both for anaphoric and exophoric uses (e.g. English *this* and *that*), which I will call demonstratives proper, and morphemes which are strictly limited to anaphoric uses (e.g. French *ce*), that is, strong articles. While the jury is still out on the question of the relative frequency of use of two morphemes in contexts where an immediately preceding clause provides an antecedent (Schwarz (2023) for first experimental evidence on English demonstratives and German strong articles), only demonstratives proper have been reported to participate in the so-called “emotional deixis” (Lakoff 1974, Potts and Schwarz 2010), illustrated in (7), and to bear stress.

(7) This cat (of mine) is driving me crazy lately.

Synchronically, we thus distinguish (at least) three types of often etymologically related morphemes: demonstratives proper, strong definite articles, and weak definite articles. If, following Bühler, we consider anaphora as a type of deixis, all three have deictic uses. Without introducing a formal machinery for capturing the distinctions in their patterning, I will nevertheless make an assumption that the relevant distinctions can be drawn in terms of the *domain* out of which a referent is supposed to be selected in each case. Specifically, I will assume that in the case of demonstratives, the referent is selected out of the most prominent situation at the time of the utterance, where the prominence equals ostentation or else defaults to the closest situation described in the discourse as most prominent. In the case of strong definites, the domain is limited to the most recent discourse situation, while for weak definites the choice of the domain is free. The diachronic chain demonstrative proper > strong definite > weak definite can in this way be associated with two seman-

tic processes: the disappearance of the prominence restriction and a further disappearance of discourse restrictions on the domain from where a referent is to be found. It is in this sense that I suggest to interpret the terms “semantic weakening” or “bleaching” as applied to the first two stages of the (enriched) Greenberg’s cycle, as schematized in Table 2.

Table 2: The first three stages of an enriched Greenberg’s cycle and associated domain restrictions

	Stage 0	Stage I	Stage II
FORM	demonstrative	strong definite	weak definite
DOMAIN	most prominent situation (default: closest discourse situation)	closest discourse situation	any situation

It needs to be stressed that the Greenberg’s cycle, is based, first and foremost, on the etymological (reconstruction) argument rather than on the observations of changing data patterns. The etymological argument rests on the fact that definite articles synchronically often correspond to phonologically weakened forms of demonstratives proper (Hawkins (2004, 83)). The argument amounts to a reconstruction of a diachronic process based on synchronic facts. In certain cases, such argument can be shown to lead to erroneous conclusions. In his discussion of the evolution of definites out of demonstratives accompanied by phonological weakening, Hawkins (2004, 83) notes that “English *the* has fewer and more reduced segments [than demonstrative *that*] (CVC, and a schwa vowel) and less stress”. Similarly, Kuteva et al. (2019, 137), citing Traugott (1980, 49), describe the following evolutionary path: “English *that*, nonproximal demonstrative > *the*, definite article”. However, in the literature focusing on English, it is generally accepted that the case-less, gender-less, and number-less *the* is a phonological continuation of the nominative masculine singular form *se*, perhaps analogically influenced by the gender and case neutral

plural form *pa*, while *that* goes back to a neutral singular accusative/nominative form *pæt* (details in section 0.3). The synchronic phonological relationship between demonstratives and definites is thus a poor diagnostic of the diachronic process. Recent advances in data annotation, search and statistical modeling make it possible to begin checking reconstruction arguments. In section 0.3 I focus on recent data-driven studies, which show that the claims based on the etymological argument, including Greenberg’s cycle, need to be substantially nuanced to match quantitative diachronic patterns.

In what follows, I discuss three case studies focusing on different stages of the enriched Greenberg’s cycle. These studies illustrate how the etymological reconstruction argument can be complemented by methodological tools that provide improved access to empirical data, which results in a better understanding of the diachronic cycles in question. The case study on the material of English, one of the most historically well-documented languages, suggests that relying on morphological shapes can be misleading and that frequencies across crucial contexts is a much better proxy to the semantics. This study also shows how surprisingly stable the underlying semantic system can be, despite morphological changes. The second and third studies revolve around the data from Latin and its descendants and provide evidence that the diachronic steps in Table 2 can indeed happen in a cyclic fashion in the sense that the reanalysis of a form from Stage X to Stage Y is associated with an emergence of a new form in Stage X (in commonly used diachronic terms, stage X gets “renewed”, e.g. Van Gelderen (2024)), but that we do not have evidence for multiple iterations of the cycle.

0.3. Case study 1: English

With the assumption that definite articles emerge as a result of diachronic changes in the semantics of demonstratives proper as a common background, approaches to the historical development of English noun phrases are divided around the question of chronology. Van Gelderen (2007), Denison (2006), Stevens (2008) propose that the Old English *se*-paradigm (Table 3) corresponds to a demonstrative, while the definite article develops in Early Middle English from the nominative form of this demonstrative (*se*). Others, such as Sommerer (2011), Crisma (2011), Allen (2016), Struik and Van Kemenade (2022), assume that the *se*-paradigm is ambiguous between demonstrative and definite interpretations already in Old English.

Table 3: Early West Saxon *se* from Hogg (1992, 143)

		SINGULAR		PLURAL
	MASC	FEM	NEUT	
NOM	<i>se, sē</i>	<i>sē</i>	<i>þæt</i>	<i>þā</i>
ACC	<i>þone</i>	<i>þā</i>	<i>þæt</i>	<i>þā</i>
GEN	<i>þæs</i>	<i>þære, þāre</i>	<i>þæs</i>	<i>þære, þāra</i>
DAT	<i>þæm, þām</i>	<i>þære, þāre</i>	<i>þæm, þām</i>	<i>þæm, þām</i>

Table 4: Old English *þes* from Mitchell and Robinson (2001, 18)

		SINGULAR		PLURAL
	MASC	FEM	NEUT	
NOM	<i>þes</i>	<i>þeos</i>	<i>þis</i>	<i>þās</i>
ACC	<i>þisne</i>	<i>þās</i>	<i>þis</i>	<i>þās</i>
GEN	<i>þisses</i>	<i>þisse, þisre</i>	<i>þisses</i>	<i>þissa, þisra</i>
DAT	<i>þissum</i>	<i>þisse</i>	<i>þissum</i>	<i>þissum</i>

Below I will discuss literature-based as well original quantitative evidence which helps settle the debate. Specifically, I will show that quantitative data allow us to dismiss the main arguments presented in the literature against assuming the existence of a definite article in the form of the unstressed *se*-paradigm in Old English. I will thus argue that in Old English we find graphically indistinguishable morphemes in Stage 0 and Stage II, and that the observed historical changes amount to the emergence of a graphically transparent morphological distinction, whereas the semantic system of determiners stays stable.

0.3.1. Interpretational uncertainty

To give some intuitive understanding of the distribution of the *se*-forms in Old English, consider a passage in (8) from a tenth-century Homilies of Ælfric, where all 5 instances are most felicitously rendered as *the* in Modern English.

- (8) He cwæð, wyt syndon an, for **ðære**₁ annysse, þæt he say.PST we be.PRS.3PL one for DEF.F.GEN oneness COMP **seo**₂ an godcundnyss, and **seo**₃ an mægenþrymnys, DEF.F.NOM one divine.nature and DEF.F.NOM one power and **þæt**₄ an gecynd þe him is and DEF.N.NOM one nature REL 3PL.MASC.DAT be.3PRS.3SG gemæne nele gepafi-an þæt he þry god-as common NOT.WANT.PRS.3SG permit-INF that they three god-PL syndon, ac an ælmihtig God æfre on ðrym had-um; and be.PRS.3PL but one almighty God ever on three person-DAT.PL and þis oncnaw-að **þa**₅ halgan þonne hi DEM.NEUT.ACC know-PRS.3PL DEF.PL saints when 3PL hinne gese-oð. 3SG.MASC.ACC see-PRS.3PL.
‘He said, We are one, because of **the**₁ unity, that **the**₂ one divine nature, and **the**₃ one mighty power, and **the**₄ one nature that is common to

them will not allow of their being three gods, but one almighty God always in three persons; and **the**₅ saints will know this when they see him.’ (Homilies of Ælfric, A supplementary Collection, ca. 1000, coaelhom,+AHom_8:194.1267, Transl. Raw (1997, 40))

Here we already encounter some of the difficulties inherent to the semantic analysis of historical languages. In particular, we must be alert to the fact that the possibility to translate a small sample of the occurrences of a morpheme *a* in language A as morpheme **b** in language B does not entail the semantic equivalence of **a** and **b**. The contexts of use of the former can turn out to be a proper subset of the contexts of use of the former in a larger sample. That is, the fact that the *se*-forms in this particular passage are felicitously translated as *the* cannot be automatically taken to mean that *se* can be modeled in exactly the same way as *the*.

In fact, the reigning uncertainty concerning the chronology of the first phase of Greenberg’s definite article cycle in English is essentially associated with the fact that in a number of cases it is possible to render a form of the *se*-paradigm as either *the* or *that*, as illustrated by the passage from Peterborough Chronicle in (9) cited by Sommerer (2011).

- (9) þa geræd-de Ulfkytel wið **þa**₁ wita-n on East
 then decide-PST.3SG Ulfcytel with DEF.PL wiseman-PL.ACC on East
 Englum. þæt him bætere weron þæt man wið
 Anglia that 3PL.DAT better be.PST.3PL COMP man with
 þone₂ here frið-es ceap-ode. ær hi
 DEF.M.ACC army.SG.ACC peace.SG.GEN buy-PST.3SG before 3PL.NOM
 to mycelne hearm on **þam**₃ earde ged-ýd-on
 too much harm.SG.ACC on DEF.M.DAT land-SG.ACC do-PST-3PL
 ‘Then Ulfcytel decided with **the**₁ councillors in east Anglia that it would
 be better that they [one] made peace with **that/the**₂ army, before they

- (11) *se* *god hat-te* *Dagon, þam* *hæðen-um*
 DEF.M.NOM god name-PST.3SG Dagon DEF.PL.DAT heathen-PL.DAT
 swyðe dyre.
 very dear
 “that god was named Dagon, very dear to the pagans” (coaelhom,+AHom 22:221.3402)

We thus have a situation where *se* can sometimes be translated only as *the*, sometimes as either *the* or *that* and sometimes only as *that*. To the extent that we maintain that *the* and *that* have different semantics in Modern English, for which we have very strong synchronic arguments, we have to conclude that *se* corresponded to two homographous paradigms, *se*₁ and *se*₂, with the semantics of *the* and *that*, respectively. Homography between definite articles and demonstratives is a common situation in contemporary Germanic languages, with however a phonological distinction of stress. For instance, in Modern Swedish the pre-nominal form *den* corresponds, if unstressed, to a definite article, in which case it combines with a suffixal determiner, or, if stressed – to a demonstrative, in which case it does not trigger the appearance of a suffixal demonstrative. The same holds for Norwegian and Danish (modulo the absence of “double definiteness” in the latter). When discussing the evolution of demonstratives into definite articles, Greenberg (1978, 62) notes that “German *der* is an unstressed variant of the demonstrative, which continues in its former use in stressed form”. This corresponds exactly to the situation I have hypothesized for Old English.

Since phonological data which would allow us to confirm the hypothesis about the co-existence of a stressed *se*-paradigm (demonstratives) and an unstressed *se*-paradigm (definite articles) in Old English are not directly accessible in historical texts, in what follows I will suggest a number of arguments

based on quantitative corpus data instead.

0.3.2. Quantitative arguments

Existing studies of the evolution of English demonstratives have been based on smaller samples than the dataset used in the present chapter, which aggregates data on noun phrases in The York-Toronto-Helsinki Parsed Corpus of Old English Prose, YCOE (Taylor et al. 2003), The Penn-Helsinki Parsed Corpus of Middle English, PPCME2 (Kroch et al. 2000), The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME, Kroch et al. 2004), and The Penn Parsed Corpus of Modern British English (PPCMBE2, Kroch et al. 2016). Figure 1 presents the data on the distribution of determiners in noun phrases in subject position excluding those with numerals, quantifiers, nominal possessives and adjectives meaning “(an)other” as potentially interfering with the use of determiners. The total number of noun phrases is 90 606. The label “that” in the legend in Figure 1 stands for *þæt*, the neutral singular form of the *se*-paradigm and its descendants in all orthographic varieties, the label “this” stands for the forms of the *þes*-paradigm (Table 4) and its historical descendant *this*, and the label “tha” for the gender- and number-neutralized form *þa*.

Figure 1 shows a number of important developments that took place between approximately 1125 and 1250. First, the nominative masculine form *se* disappears.⁴ Second, the frequency of the gender-neutralized form *þa* dramatically decreases, and in the next period it shares the fate of *se*. Third, the frequency of the gender-, number-, and case- neutralized form *the* skyrockets. Fourth, over a larger timespan, we can observe that as the frequency of *the*

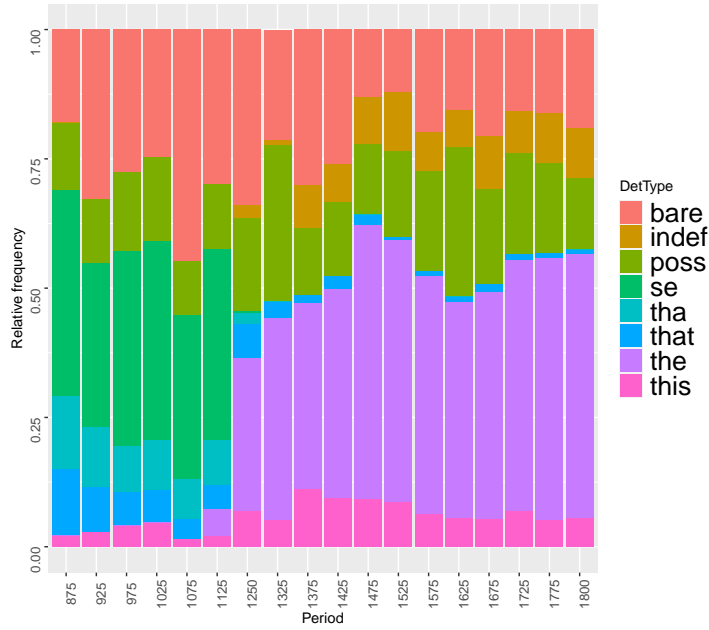


Figure 1: Determiner types in subject noun phrases in English

grows, the frequency of *pæt/that* goes down. Finally, around 1125 and 1225, the indefinite article enters the determiner system.

In the debate about whether Old English *se* functioned as a demonstrative or a definite article, occurrences of bare noun phrases in contexts which require the use of *the* in PDE play an important role. An example is provided in (12). Such cases can be taken as evidence in favor of a demonstrative semantics of the *se* paradigm in Old English.

- (12) Her sunne aðeostrode on xii kl. Iulii
 here sun grew.dark on 20 July
 “Here **the** sun grew dark on June 20.” (cochronE,ChronE-[Plummer]:540.1.183)S23,
 cited from Stevens (2008)

While the argument of “non-obligatoriness” of the *se*-forms is brought up in

virtually all discussions of the Old English grammar, in quantitative terms its importance is of marginal or even negligible significance. In addition to large scale corpus data to be presented shortly, crucial evidence for this conclusion comes from relatively small samples carefully hand-annotated for the relevant pragmatic feature, namely, whether the conditions triggering the appearance of *the* in Modern English are satisfied. Specifically, Sommerer (2011, 214) shows that in a sample from *Parker Chronicle* (ca. 1000), among 824 noun phrases, which, if translated into Modern English, would require an overt marker of definiteness (a definite article, a demonstrative, or a possessive pronoun), there are 8 bare NPs.⁵ That is, only 1% of noun phrases requiring a definite interpretation do not feature a definiteness marker. Sommerer (2011, 214) obtains the same result for a sample from *Peterborough Chronicle* (ca. 1200), where she finds 32 bare noun phrases among 3105 noun phrases requiring a definite interpretation, which, again, equals 1% of occurrences of zero determination. Crisma (2011) comes to a similar conclusion based on examining the ninth century prose, where she does not find instances of bare noun phrases interpreted definitely. To this, I adduce the results of a study of the use of determiners in superlative noun phrases, which are morphologically conditioned to require a definite interpretation. By virtue of the semantics of the superlative morpheme, such noun phrases necessarily denote a unique individual possessing the quality described by the NP to the highest degree. Among 551 superlative noun phrases in the corpora, there are 14 ($\approx 2\%$) of bare noun phrases.

Evaluating the frequency of definiteness marking in much larger samples is impractical as it necessitates identifying all noun phrases requiring the presence of a definiteness marker, which requires an unambiguous understanding of the relevant context not always feasible for historical texts. Instead, in larger

samples we can easily track the frequency of definiteness markers. I do this by fitting a logistic regression model in (13) to the binary variable absence/presence of a definiteness marker, which includes the forms suspected of having either a definite or a demonstrative semantics (*se*-paradigm forms, *bes*-paradigm forms, *the*, and possessive pronouns).

$$(13) \quad P(\text{DEF} = \text{yes} | \text{DATE} = d) = \frac{e^{\alpha + \beta * \text{Date}}}{1 + e^{\alpha + \beta * \text{Date}}}$$

The curve of the model is plotted in Figure 2. The coefficient of the Date reflecting the impact of time on the frequency of definiteness markers is estimated to be very small ($\beta = 0.00004$) and marginally significant ($p = 0.036$), meaning that there is very little diachronic change.

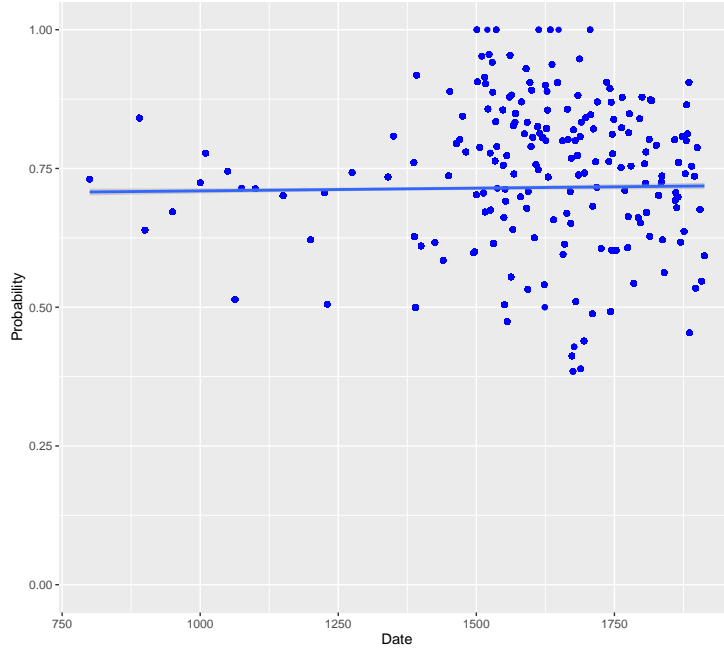


Figure 2: Presence of a definiteness marker in subject noun phrases in English, logistic regression

The fact that the overall frequency of definiteness markers does not appear to change very much across time in English in and of itself does not tell us anything about the semantics of particular paradigms involved. In particular, it does not clarify whether the *se*-paradigm had the semantics of a demonstrative or of a definite article. In the absence of phonological information, it is often impossible to determine what the intended interpretation of a given instance of a *se*-form was (cf. the discussion of ambiguously translatable cases above). In such cases, typological information can come to the rescue, especially when combined with the assumption that the pragmatic conditions governing the use of demonstratives, as well as the frequency with which these conditions are met by the Common Ground, remain constant over time. This assumption can be considered part of the Uniform Probabilities Principle (Lass 1997, 28) which states that “[t]he (global, cross-linguistic) likelihood of any linguistic state of affairs (structure, inventory, process, etc.) has always been roughly the same as it is now” (on Uniformitarianism see WBCDL008). Table 5 gives the frequency of proximal and distal demonstratives per 100 words in adult speech across a sample of languages, which can be compared with the proportions of *se*- and *þes*-paradigm forms per 100 words in Old English texts binned into 50-year intervals in Table 6.

Table 5: Mean proportions of demonstratives per 100 words in adults’ speech, as reported by Diessel and Monakhov (2023, 936)

	Proximal mean	Distal mean
English	1.89	4.91
French	0.11	1.53
Spanish	2.08	0.13
Japanese	3.74	0.36
Chinese	2.13	0.76
Hebrew	7.43	0.31

Table 6: Proportions of *se*- and *bes*-paradigm forms per 100 words in Old English texts (per 50-year period)

	<i>bes</i> -paradigm	<i>se</i> -paradigm
875	1.55	7.18
925	4.32	17.13
975	7.89	23.9
1025	13.2	36.6
1075	4.49	18.03

While varying from period to period, likely due to extralinguistic factors such as the overrepresentation of a particular text genre in a given time slot, the proportions of the *bes*-paradigm forms stay close to the typological range. The relative frequency of the *se*-paradigm forms, in contrast, is one or even two orders of magnitude higher than proportions attested in the sample provided by Diessel and Monakhov (2023). Unless we assume that the conditions for using demonstratives were satisfied in Old English texts 10 to 100 times more often than in the languages listed in Table 5, which does not seem likely, the most straightforward explanation for this observation is that the *se*-paradigm forms in Old English corresponds to a *mixture* of morphemes with a definite

article semantics and (homographous) morphemes with a properly demonstrative semantics. I conclude, therefore, that during the Late Old English period, or roughly until 1150, the *se*-forms (including *se*, *þæt*, and *þa*) corresponded to a definite article analogous to Modern English *the*, and to a demonstrative, the two being distinguishable phonologically but not graphically. After this period, the homography disappears and gender-, case- and number-neutralized form *the* replaces all instances of the *se*-paradigm on the properly definite interpretation, including those involving neutral singular nouns previously covered by *þæt/that*. This hypothesis is further corroborated by the sudden drop in the frequency of *þæt/that* observable in Figure 1. This drop is expected if *þæt/that* ceases to graphically correspond to both a definite article and a demonstrative, and only corresponds to a gender, case-, and number neutralized demonstrative.

The case of Old and Middle English testifies to a greater diachronic stability of the system of definiteness marking than previously thought. *Modulo* the disappearance or reduction of gender-, case-, and number-agreement, the system of definiteness markers can be argued to have remained intact since at least Late Old to Modern English, with a dedicated set of morphemes used as demonstratives and another as definite articles. With regard to Greenberg’s cycle, there are no grounds to assume a shift from the semantics of a demonstrative proper to the semantics of a strong or weak definite in the period from Late Old English to Modern English.

The shift that appears to be empirically unobservable in English, seems to have occurred within the available diachronic window in Northern Germanic languages, where the suffixal definite article is commonly hypothesized to originate from a free-standing demonstrative *hinn* “yon”. Piotrowska and Skrzypek

(2021) show, on the basis of a corpus of Danish, Old Icelandic, and Old Swedish texts from 1200–1550, that the probability of the use of the suffixal article, as estimated by a logistic regression model, rises at a statistically significant rate across this period (the time coefficient $\beta=0.004$, $p = 0.0001$). Comparing this with the rate of change I estimated for English definiteness markers above ($\beta=0.00004$), the Mainland Scandinavian rate of change is 100 times higher, suggesting an actual diachronic development in the latter case. Conceptually, it is far from self-evident what it means that the frequency of a morpheme grows with time. Simonenko and Carlier (2020) propose that the French article *le*, which, similarly to the Mainland Scandinavian suffix, grows in frequency at a non-negligible rate during the course of Old and Middle French periods, is during this time ambiguous between a strong definite and a weak definite interpretations, the weak definite interpretation becoming more frequent with time. That it could sometimes be interpreted as a strong definite accounts for the (diachronically decreasing) share of bare noun phrases which would require *le* in today’s French. In this case, as in Old English and in Old Mainland Scandinavian, we do not have empirical access to a stage whereby the morpheme in question would be in Stage 0 or (only) in Stage I (strong definite).

A stage where a Germanic *sa* paradigm has not yet shifted to Stage II (that is, weak definite), appears to be attested for Gothic. Giurgea (2024) reports, on the basis of comparing Gothic biblical translations attributed to the first Gothic bishop Wulfila (c. 307–383) with the Greek original, that the *sa*-paradigm is used almost exclusively in the context of anaphoric and exophoric reference, but not in other types of definite contexts. Out of 230 noun phrases which have an antecedent in the preceding (local) context, 218 (95%) feature a form from the *sa*-paradigm. In contrast, among NPs without an antecedent

but where the nominal predicate denotes a maximal individual in a particular situation, a *sa*-paradigm form appears only in 4% (13 out of 322 NPs), the rest being bare NPs. It is clear that in Gothic and in Old English, where the *se*-paradigm occurs in about 98%-99% of NPs satisfying the maximality condition, *s*-paradigms have completely different distributions. This fact should be reflected in different semantic representations.

In the next two sections I discuss how the evolution of individual items relates to the evolution of the determiner system as a whole, using Romance data for two case studies, which attest to the chain-shift nature of the evolutionary developments which start with demonstratives proper.

0.4. Paradigms and chain shifts

Demonstratives are sometimes categorized as “personal-spatial” deictics (Ledgeway and Smith 2016), since their meaning incorporates reference to the position of speech participants. Brugmann (1904) first proposed to distinguish four *Zeigarten* or modes of pointing defined in terms of the spheres of the speech act participants: *Der*-deixis, *Ich*-deixis, *Du*-deixis, and *Jener*-deixis. Brugmann’s classification was further developed (and relabeled) in Wackernagel (1918), who gives the definitions of the four types as in (14)–(17).

- (14) (Greek) *τό*-deixis, “in which no distinction is made between the near and the remote, but a rather vague indication is given of something not immediately adjacent to the speaker”;
- (15) (Latin) *hic*-deixis (Greek *ὅδε*), “when the pointing is at the speaker himself, and the addressee is enjoined to direct his gaze at the ‘I’ of

the speaker, at his sphere, at what must lie closest to his circle of thought”;

- (16) (Latin) *iste*-deixis (Greek *οὗτος*), “whereby the addressee’s attention is directed at something confronting the speaker”;

- (17) (Latin) *ille*-deixis (Greek *ἐκεῖνος*), “which serves to indicate something more remote in space or time, or something beyond a certain limit”.

As suggested by Wackernagel’s choice of a mixture of Greek and Latin morphemes as labels of different *Zeigarten*, he was not aware of a language maintaining the four-way distinction. Wackernagel notes that Latin lacks a correlate of the Greek *τό*-deixis, Sanskrit and Old Iranian do not distinguish *iste*-deixis and Osco-Umbrian, despite being in other respects very closely related to Latin, lacks the *hic* and *ille* paradigms. Wackernagel (1918) further notes that the way the morphological inventory partitions the contexts of use can differ from language to language. As an example, he mentions that “Lat[in] *iste* has a much stronger connection with the addressee (2nd pers.) than G[ree]k *οὗτος*), which sometimes even refers to the person of the speaker”. Personal-spatial deictics were thus from early on considered as a system of meaning-generating oppositions.

Frei (1944) develops this insight further, arguing that the interpretation of a given demonstrative form is a function of the number of oppositions in the demonstrative system, first classification of (Romance) demonstrative systems into binary and ternary ones going back to Meyer-Lübke (1895). Frei (1944) argues that the so-called proximal members of the binary systems are used to relate the denotation to (the sphere of) the speaker or the hearer, whereas the first member of the ternary oppositions can only serve to relate the denotation

to the sphere of the speaker. In particular, he attributes the difference in the interpretation between French *celui-ci*, German *dieser*, and English *this* on the one hand and Latin *hic*, Greek *ὁδε*, Old Slavic *сь* and Armenian *ays* on the other to the fact that the former are part of binary systems, while the latter are members of ternary oppositions (*celui-ci/celui-là*, *dieser/jener*, *this/that* vs. *hic/iste/illic*, *ὁδε/οὐτοοσ/ἐκεῖνοσ*, *сь/тъ/онъ*, *ays/ayd/ayn*).

According to Diessel (2012b, 3-4), the interpretation of individual demonstrative forms is a function of both the architecture of the system and the current context, whereby “*here* and *there*, and other proximal and distal deictics, do not express absolute measure of deixis, but differentiate between two different locations relative to the deictic centre within the current construal of the speech”, *here* denoting the “area that is conceptualized as the deictic centre, and *there* indicat[ing] a location that is not included in this area”. Levinson (2004) formulates a similar insight in neo-Gricean terms, arguing that the distal interpretation of the semantically neutral *that* in English is a scalar implicature arising in competition with the (inherently proximal) *this*.

The question about semantic and cognitive correlates of the formal distinctions within demonstrative systems is however far from settled. Kemmerer (1999) argues that while our perceptual system incorporates a sharp neurologically-grounded distinction between the sphere within one’s arm’s reach and the sphere beyond that, the patterning of proximal vs. distal demonstratives cuts across this distinction and has to appeal to other categories. A series of more recent studies, including experimental ones, shifts from defining proximal/distal features in terms of the physical or mental spaces associated with the speech act participants towards the notions that have to do with the flow of information in the discourse (e.g. references in Næss et al. (2020, 1), Piwek et al.

(2008), Rocca and Wallentin (2020), Coventry et al. (2023)).

Piwek et al. (2008, 696), in particular, hypothesize that “proximals are used for intense indicating and distals for neutral indicating”, where intensity inversely correlates with the referent’s accessibility, which, in turn, is a function of being in the focus of attention. A referent lies within the focus of attention and therefore has high accessibility if either 1) it “was referred to in the preceding utterance or is adjacent to an object that was referred to in the preceding utterance” or 2) it “lies in an area to which the speaker explicitly directed the attention of the addressee”. The hypothesis predicts that distal demonstratives (neutral indicating), such as English *that* or Latin *ille*, are used to refer to entities which already lie within the focus of attention, whereas proximals (intense indicating), such as English *this* or Latin *hic* and *iste* encode newly focused information.

Typologically, most frequent demonstrative systems include two or three terms. Estimates made by Diessel (2013) on a sample of 234 languages, suggest that 54% of languages have a demonstrative system with a binary opposition, 37% ternary, about 3% maintain a four-way contrast and 2% feature a single-membered system, and the rest five- or more membered systems.

With regard to diachronic trends within demonstrative system, Frei (1944) makes the first, to my knowledge, generalization, arguing that the number of oppositions tends to decrease, rather than increase, with time. Coupled with the assumption that the meaning of demonstratives depends on the structure of the demonstrative system as a whole, this generalization predicts that the remaining members of the reduced system will universally shift their meaning. In particular, it is expected that one or both members of the new binary system will expand their use to cover the distributional area of the term that fell out

of use. I formulate this expectation as “Frei’s conjecture” in (18).

- (18) FREI’S CONJECTURE: The diachronic reduction in the morphological inventory of demonstratives is accompanied by the expansion of the distribution of one or more of (the reflexes of) the members.

Frei’s conjecture establishes a bridge between two evolutionary dimensions: changes in the architecture of personal-spatial systems are predicted to go hand in hand with changes in the semantics of individual demonstratives. In particular, the following empirical observation falls out of the conjecture. It has been observed that if a language develops a specialized anaphoric demonstrative (a strong definite article in the terminology used in this chapter), most often it is a formerly distal member of the proximal-distal demonstrative opposition (Christophersen 1939). That is, a semantic change is accompanied by the expansion of the use of the distal form, whereby the distal/proximal opposition is neutralized and a form that could have been used exophorically and emphatically turns into a strong definite article.

0.5. Case study II: Latin

In addition to the ternary system proper demonstratives *hic/iste/il**le*, Classical Latin features a strictly anaphoric demonstrative (i.e. strong definite) *is*, as well as form traditionally considered as a focus demonstrative, *ipse* “himself/herself”. While in Classical Latin both *ille* and *is* can be used anaphorically to refer to a previously mentioned referent, *ille* in subject position often indicates a topic shift and *is* indicates topic continuity (Danckaert et al. 2021, on *is* as a short-distance anaphor see also André and Fruyt (2012), Guillot-Barbance

and Marchello-Nizia (2015)). The changes of the demonstrative system from Classical to Late Latin are primarily characterized by *ille* encroaching on the distribution of *is* (Carlier (2017), Carlier and Guillot-Barbance (2018)). The replacement is illustrated in Figure 3 for adnominal uses (Late Latin data are provided by Anne Carlier, Classical Latin counts come from the treebanks of the projects Crane (Perseus) and PROIEL by Haug and Johndal (2008)) and pronominal uses (plotted from the counts in Danckaert et al. (2021)). These corpus results corroborate the conclusion reached by Vänänen (1981, 121) and André and Fruyt (2012), among others, about *ille* replacing *is* in Latin.

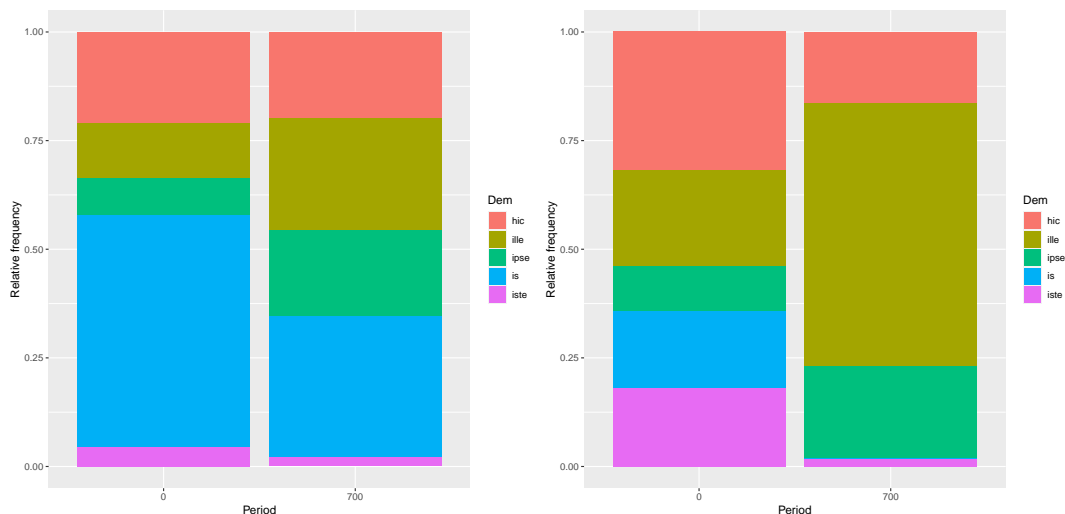


Figure 3: Frequencies of adnominal (left) and pronominal (right) demonstratives in Classical and Late Latin

The quantitative replacement can be taken as indicative of the evolution of the meaning of *ille* itself from a demonstrative proper to a strong definite marking short-distance anaphoric relations. Guillot-Barbance and Marchello-Nizia (2015)), in particular, note that in earlier texts *ille* appears in cases where

the relevant referent was introduced a relatively long time ago, while in later texts *ille* functions as a short-distance anaphor. The results of studies of relative frequencies of *ille* and *is* thus converge with the results of an examination of the changes in the length of anaphoric dependencies for *ille* on the hypothesis that it shifted from Stage 0 to Stage I of the enriched Greenberg’s cycle. A study formally and systematically measuring the evolution of the anaphoric distance in the uses of *ille* still needs to be done.

The passage of *ille* from Stage 0 to Stage I has been related to the spread of new formations containing the interjection *ecce* “Here is!” (<eccu<*akke/*akkū, Carlier and De Mulder (2010), Ledgeway (2020)). These forms are reflected in most Romance languages as demonstratives proper. In the next section, I examine their fate in French, where they undergo the first Stage 0 → Stage I shift of the enriched Greenberg’s cycle becoming a strong definite, which, in turn, has also been related to the emergence of the next generation of demonstratives proper.

0.6. Case study III: French

The Latin system of adnominal demonstratives is reflected in the earliest attestations of Old French as proximal/distal paradigms *cist/cil* and *le*-paradigm (reflex of *ille*). In addition, Old and Middle French feature a number of innovative forms: a paradigm *ce*, taken to result from the proximal/distal neutralization of *cist/cil* (Marchello-Nizia 2005), paradigms with a prefix *i-* *icist/icil* and postnominal locative particles *-ci* and *-là* which combine with various adnominal demonstratives. This is summarized in Table 7, together with the dates for the first attestations in the corpus of Kroch and Santorini (2021).

Table 7: French demonstratives: Latin descendants and innovations

Latin etymon	French innovation	First corpus attestation
ecce + istum	(c)ist	842 (first corpus text)
	icist	1050
	cist -ci	1279
	cist -là	1498
ecce + illum	cil	900
	icil	1050
	cil -là	1225
	cil -ci	1373
	ce	1000
	ice	1050
	ce -là	1156
	ce -ci	1330
ille	le	900

Reflecting a general lack of consensus with regard to the semantics of the proximal/distal distinction, the contrast between French *cist* and *cil* has been modeled in terms of a) reference to an individual within vs. outside of the situation of the utterance (Guiraud 1967), b) a referent being identifiable vs. non-identifiable by means of (intra- or extra linguistic) material in the immediate vicinity of the occurrence of the demonstrative (Kleiber 1987), and, finally, c) reference to an individual located within vs. outside of the (subjectively defined) sphere of the speaker (Marchello-Nizia 2003, Guillot-Barbance 2017).

The chronological order of the first attestations of the circumfix-like formations matches the observations made by Guillot-Barbance and Marchello-Nizia (2015, 98) that the appearance of *cil N-là* and *cist N-ci* predates that of *cil N-ci* and *cist N-là*. Assuming that initially *cist* and *cil* form a semantic oppo-

sition along the proximal/distal axis (defined, following Marchello-Nizia 2003, as reference within/outside the sphere of the speaker), and that *ci* and *là* form a parallel opposition, the appearance of formations with both (etymologically) proximal and distal components, such as *cil-ci*, can plausibly be taken to indicate a neutralization of the proximal/distal contrast between *cist* and *cil*. As mentioned in section 0.2, I assume that the demonstrative proper semantic involves picking a referent out of the most prominent situation by means of ostentation or by defaulting to the anaphorically closest (and thus most prominent) situation, while in the case of strong definites the reference domain is limited to the anaphorically closest situation. On this view, the proximal/distal distinction can be indicative of the demonstrative proper semantics, as it presumably provides additional restrictions on the prominent situation choice (e.g. speaker’s sphere). If so, the proximal/distal neutralization can be seen as a hallmark of the shift from Stage 0 to Stage I (i.e. from a demonstrative to a strong definite semantics).

The forms prefixed with *i-* – *ice*, *icist*, and *icil* are analyzed by Guillot-Barbance and Marchello-Nizia (2015, 90) in terms of “re-déixisation” of *cist* and *cil*. I interpret this in terms of the emergence of new or renewed Stage 0 forms (demonstratives proper) once *cist* and *cil* become strong definites. According to De Mulder et al. (2010), the prefixed forms *icist* and *icil* were used for focalization and were stressed. On their analysis, the French phonological system undergoing a transition from word-level to syntagmatic stress caused the stressed pre-nominal demonstratives to be replaced with circumfix-like formations where the new exophoric element occurs in the final position in the nominal syntagma and thus can receive prosodic prominence (i.e. *icil chevalier* → *ce chevalier-là*), while *ce* stays unaccented (Marchello-Nizia 1995).

In Modern French, as pointed out already by Frei (1944), the opposition of *ce* NP-*ci* and *ce* NP-*là* is disappearing, with *ce* NP-*là* expanding onto the “original” distributional area of *ce* NP-*ci*.

Figure 4 shows the quantitative diachronic distribution of various adnominal demonstratives (total 10600) based on the corpus of Kroch and Santorini (2021).

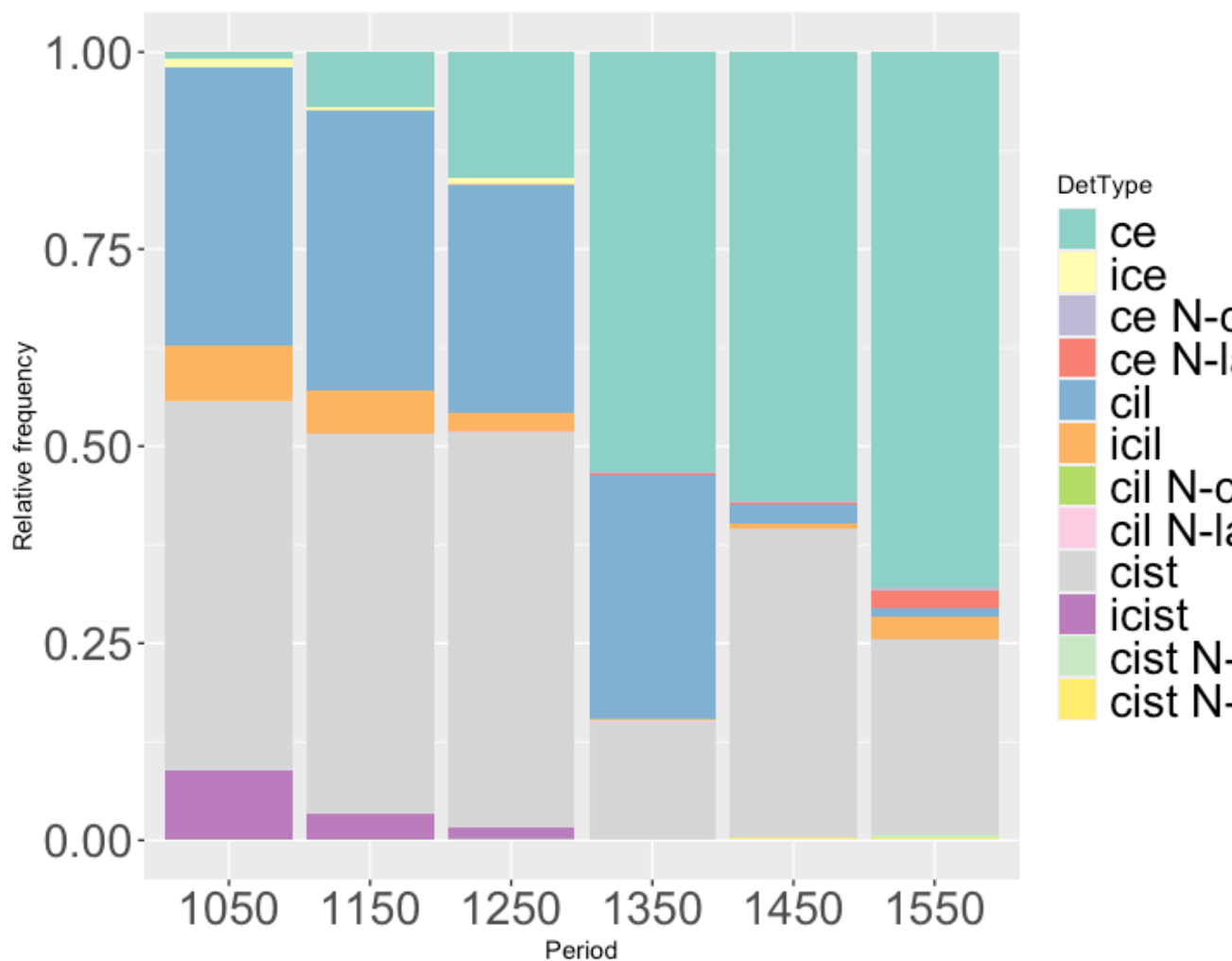


Figure 4: Frequencies of adnominal demonstratives in French

The quantitative picture introduces complications into the neat hypothetical timeline whereby as *cil* and *cist* are replaced by *ce* as they shift from Stage 0 to Stage I, with Stage 0 being renewed first by prefixed and then by circumfixed formations. In particular, based on the quantitative results we cannot say that the prefixed formations replace *cil* and *cist*: for the first three periods the relative frequency of the prefixed formations does not grow (if anything, it declines), and then they virtually disappear without being themselves replaced in terms of relative frequencies. One possible scenario compatible with the data is that *cil* and *cist* for a long period of time were alternating between Stage 0 and Stage I interpretations (demonstrative proper vs. strong definite), with the Stage 1 uses being progressively replaced by *ce*. The remaining possibility for *cil* and *cist* to be interpreted as demonstratives proper accounts for why we do not see a surge in prefixed or circumfixed formation. While I will not further develop this possibility here, it is clear that to fully understand the picture, the enriched Greenberg’s cycle of change needs to be supplemented with a variation model of language change.

It has first been observed by Pedersen (1905) (cited from Frei (1944)) that the Classical Latin ternary system of (exophoric) demonstratives *hic/iste/ille* in some Romance varieties was transformed into a binary system, while in others it is reflected as a ternary system. Ledgeway (2020) gives a summary, replicated in Table 8, of the modern Romance binary systems where the members of the pair are descendants of *ecce ill(um)* and *ecce ist(um)*. As predicted by Frei’s Conjecture in (18), the descendants of Latin hearer-oriented *iste* are not restricted to the uses associated with the hearer in these binary systems, but rather cover all uses where referents are “subjectively perceived to fall within the deictic centre” (Ledgeway 2020, 453).⁶

Table 8: Romance varieties with a binary demonstrative system from Ledgeway (2020)

reflex of	ECCE IST(UM)	ECCE ILL(UM)
Occitan	<i>aqueste</i>	<i>aquel/aquéu</i>
Gascon (Testerin)	<i>aquis</i>	<i>aquits</i>
Ladin	<i>chësc</i>	<i>chël</i>
Northern Italian dialects	<i>(cu)st</i>	<i>cul</i>
Italian	<i>questo</i>	<i>quello</i>
Vegliot	<i>kost</i>	<i>kol</i>
Romanian	<i>acesta</i>	<i>acela</i>
Southern Daco-Romance	<i>ăsta</i>	<i>ăla</i>
Moldovan	<i>aista</i>	<i>acela</i>
Megleno-Romance	<i>tsista</i>	<i>tsela</i>

Ledgeway (2020) hypothesizes a historical transitional stage whereby reflexes of *iste* “marked the shared deictic spheres of both discourse participants” and gives as examples Old French and Raeto-Romance varieties Surselvan and Vallader (based on the data in Sornicola (2011)). The deictic centre during such stages for him necessarily involves both the speaker and the hearer. In such a system it is presumably not possible to use a reflex of *iste* to refer to an entity which clearly does not belong to the sphere of the hearer. Carlier (2017), following Marchello-Nizia (2005), argues that depending on the context, the deictic centre Latin could involve either the speaker or the speaker and the hearer together. In the latter case, in Latin either *hic* or *iste* could have been used, which made it possible to reanalyse *iste* as speaker-oriented. On this view, Old French *iste* just inherits the two innovative uses of Latin *iste*: as alluding to the deictic centre as the sphere of the speaker and the hearer together, or as the speaker only. It remains to be seen if these hypotheses can be formally tested on the historical textual data, often compatible with more

than one reading, and given the challenges associated with probing into the semantics of the proximal/distal distinction in present-day varieties even in experimental settings (see the discussion in section 0.4).

Ledgeway (2020, 455) indicates that the varieties listed in Table 8 frequently make use of formations involving locative particles meaning ‘here’ and ‘there’, which, “although originally emphatic in nature, are today generally unmarked and often preferred”.

Some other Romance varieties, such as Northern Italian dialects Ligurian, (19), and Friulian, (20), feature a complete neutralization of the proximal/distal opposition between the reflexes of *istum* and *illum*, as attested by the “mismatched” forms, where the etymologically proximal or distal member of a demonstrative proper opposition becomes a form which on its patterns as a strong definite form and can be combined with a locative particle to form a “new” demonstrative proper. (cf. Old French *cil-ci* and *cist-lá*, Table 7).

(19) *stu ki invece de stu là*
 this here instead of this there
 “This one instead of that one” Ligurian (Azaretti (1982), cited from
 Ledgeway (2020, 456))

(20) *kel libri ka/la*
 that book here/there
 “This/That book” Friulian (Vanelli (1997), cited from Ledgeway (2020,
 456))

These facts point to the “renewal” of Stage 0 semantics of the Latin reflexes and their subsequent shift to Stage I of the enriched Greenberg’s cycle. Given the current absence of large annotated historical corpora for most of these varieties, establishing the timeframe of these processes remains a task for the

future.

0.7. Conclusions

In this chapter I have discussed theoretical issues related to historical developments of adnominal demonstratives against the backdrop of a methodological paradigm transition from the so-called etymological argumentation to arguments based on the observation of changing quantitative patterns in better documented languages. The methodological shift challenges some of the previously held empirical assumptions and calls for new theoretical answers.

In particular, I have shown that quantitative evidence for Old and Middle English challenges the long-held view of a historically “observable” shift from a demonstrative to a definite article paradigm. Among other things, this conclusion suggests a greater diachronic stability of demonstrative systems than was assumed based on the etymological arguments. This conclusion is further strengthened by case studies on Latin and French. Based on the available historical data, we observe a shift from Stage 0 to Stage I happening once in Latin, where a demonstrative proper *ille* becomes a strong definite with short anaphoric chains, and once in French, where *cil/cist* lose their demonstrative proper semantics (and dectic opposition) and are replaced by a strong definite *ce*. Strictly speaking, in both cases, *within a language* (how to define it historically being of course a non-trivial question), we do not have grounds to talk about a cycle in the sense of multiple instances of Stage 0 to Stage 1 shift. The “renewal” of Stage is observed in Latin also only once, with formations involving *ecce* beginning to accompany the “old” demonstratives. In French, the particles *là* and *ci* that “renewed” Stage 0 after or in parallel with the demise

of *cist/cil* are used up to this day, with *ce N ci-bas* and *ce N là-bas* replacing *ce N-ci* and *ce N-là* in Present Day French. The case studies of English, Latin, and French thus suggest that diachronic processes affecting demonstratives systems are extremely slow, to the point that we may tentatively hypothesize that every shift in Greenberg’s “cycle” appears only once in one language, further iterations affecting reflexes in a system that has overall changed so much that we refer to it as another language. Putting the hypothesis to test requires a larger diachronic sample.

As the quantitative study of historical French suggests (Figure 4), as well as the discussion of the Romance patterns in general in Ledgeway (2020), the neat distinction between various n-ary demonstrative systems is to a large extent an idealization, since the number of oppositions and their morphological realization changes gradually, with forms maintaining a “double identity” which can only be captured within a variationist approach to language change.

Bibliography

- Allen, C. L. (2016). The definite determiner in Early Middle English: What happened with *þe*? In *Let us have articles betwixt us: Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood*. Dept. of English, School of Communication & Culture, Aarhus University.
- André, A. and M. Fruyt (2012). Le rôle de *is* dans les changements de l’endophore et de la déixis en latin. *MOM Éditions* 49(1), 291–307.
- Azaretti, E. (1982). *L’evoluzione dei dialetti liguri*. Sanremo: Casabianca.
- Benveniste, E. (1956). The nature of pronouns. In M. Halle, H. G. Lunt,

- H. McLean, and C. H. van Schooneveld (Eds.), *For Roman Jakobson*, pp. 34–37. The Hague: Mouton.
- Brugmann, K. (1904). Die Demonstrativpronomen der indogermanischen Sprachen. Eine bedeutungsgeschichtliche Untersuchung. *Abhandlungen der philologisch-historischen Klasse der Königl. Sächsischen Gesellschaft der Wissenschaften* 22(6).
- Bühler, K. (1934). *Theory of Language: The representational function of language* [Goodwin, Donald Fraser, trans. 1990]. Amsterdam: John Benjamins.
- Buridant, C. (2000). *Grammaire nouvelle de l'ancien français*. Paris: Sedes.
- Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The New Psychology of Language*, pp. 145–168. Mahwah NJ: Laurence Erlbaum Associates.
- Carlier, A. (2017). Le système des démonstratifs en cours de restructuration en latin tardif: une séparation des rôles référentiel et pragmatique de la deixis. *Langages* 208(4), 29–52.
- Carlier, A. and W. De Mulder (2010). The emergence of the definite article in Late Latin: *ille* in competition with *ipse*. In H. Cuykens, K. Davidse, and L. Van de Lanotte (Eds.), *Subjectification, Intersubjectification and Grammaticalization*, pp. 241–275. The Hague: Mouton De Gruyter.
- Carlier, A. and C. Guillot-Barbance (2018). The restructuring of the demonstrative paradigm in the transition from latin to french. In A. Carlier and C. Guillot-Barbance (Eds.), *Latin tardif, français ancien: continuités et ruptures*, Volume 420, pp. 201–232. Walter de Gruyter GmbH & Co KG.

- Christophersen, P. (1939). *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.
- Coventry, K. R., H. B. Gudde, H. Diessel, J. Collier, P. Guijarro-Fuentes, M. Vulchanova, V. Vulchanov, E. Todisco, M. Reile, M. Breunesse, et al. (2023). Spatial communication systems across languages reflect universal action constraints. *Nature human behaviour* 7(12), 2099–2110.
- Crisma, P. (2011). The emergence of the definite article in English. In P. Sleeman and H. Perridon (Eds.), *The Noun Phrase in Romance and Germanic. Structure, variation, and change*, pp. 175–192. Amsterdam: John Benjamins.
- Danckaert, L., L. Haegeman, and S. Prévost (2021). How lexical merger can drive grammaticalization. *Journal of Historical Syntax* 5(16-25), 1–34.
- De Mulder, W., C. Guillot, and J. Mortelmans (2010). *Ce N-ci et ce N-là en moyen français*. In L. M. Tovenia (Ed.), *Déterminants en diachronie et synchronie*, pp. 86–103. Paris: Projet ELICO Publications.
- Denison, D. (2006). Category change and gradience in the determiner system. In B. L. Ans van Kemenade (Ed.), *The handbook of the history of English*, pp. 279–304. Wiley Online Library.
- Diessel, H. (1999). The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3, 1–49.
- Diessel, H. (2012a). Bühler’s two-field theory of pointing and naming and the deictic origins of grammatical morphemes. In T. Breban, L. Brems, K. Davidse, and T. Mortelmans (Eds.), *New Perspectives on Grammaticalization: Theoretical Understanding and Empirical Description*, pp. 37–50. Amsterdam: John Benjamins.

- Diessel, H. (2012b). Deixis and demonstratives. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics*, pp. 1–25. Berlin: Mouton De Gruyter.
- Diessel, H. (2013). Distance contrasts in demonstratives (v2020.3). In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Zenodo.
- Diessel, H. and S. Monakhov (2023). Acquisition of demonstratives in cross-linguistic perspective. *Journal of Child Language* 50(4), 922–953.
- Ebert, K. (1970). *Referenz, Sprechsituation und die bestimmte Artikel in einem nordfriesischen Dialekt*. Ph. D. thesis, Nordfriisk Institut, Bräist/Bredstedt.
- Eckermann, W. (2007). *A descriptive grammar of the Bukawa language of the Morobe Province of Papua New Guinea*. Pacific Linguistics.
- Fillmore, C. J. (1971). Santa cruz lectures on deixis. [published in 1997 as lectures in deixis].
- Fraurud, K. (2001). Possessives with extensive use. In I. Baron, F. Sørensen, and M. Herslund (Eds.), *Dimensions of Possession: a source of definite articles*, pp. 243–267. John Benjamins.
- Frei, H. (1944). Systèmes de déictiques. *Acta linguistica* 4(1), 111–129.
- Fuchs, M. and P. B. Schumacher (2020). Referential shift potential of demonstrative pronouns – evidence from text continuation. In Å. Næss, A. Margetts, and Y. Treis (Eds.), *Demonstratives in discourse*, pp. 185–213. Language Science Press.

- Giurgea, I. (2012). The origin of the Romanian “possessive-genitival article” *al* and the development of the demonstrative system. *Revue Roumaine de Linguistique LVII*(1), 35–65.
- Giurgea, I. (2024). A case study in the development of definite articles: The “anaphoric article” of Gothic. Manuscript, the “Iorgu Iordan-Al.Rosetti” Institute of Linguistics of the Romanian Academy, Bucharest.
- Greenberg, J. H. (1978). How does a language acquire gender markers? *Universals of Human Language 3*, 47–82.
- Guillot-Barbance, C. (2017). *Le démonstratif en français: étude de sémantique grammaticale diachronique (9ème-15ème siècles)*. Leuven: Peeters.
- Guillot-Barbance, C. and C. Marchello-Nizia (2015). Spécialisation morpho-syntaxique et changement sémantique: le cas du démonstratif français. *Langue française 187*, 79–110.
- Guiraud, P. (1967). L’assiette du nom dans La Chanson de Roland. *Romania 88*, 59–83.
- Haug, D. T. T. and M. L. Johndal (2008). Creating a parallel treebank of the old indo-european bible translations. In C. Sporleder and K. Ribarov (Eds.), *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, pp. 27–34.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: OUP.
- Heine, B. and T. Kuteva (2007). *The genesis of grammar: A reconstruction*. Oxford: OUP.

- Hogg, R. M. (1992). *The Cambridge history of the English language*, Volume 1. Cambridge: Cambridge University Press.
- Kemmerer, D. (1999). “near” and “far” in language and perception. *Cognition* 73(1), 35–63.
- Kleiber, G. (1987). L’opposition *cist/cil* en ancien français ou comment analyser les démonstratifs? *Revue de linguistique romane* 51(201–202), 5–35.
- Kleiber, G. (1990). *La Deixis*. Paris: Presses Universitaires de France.
- Kroch, A. and B. Santorini (2021). *Penn-BFM Parsed Corpus of Historical French (PPCHF)*.
- Kroch, A., B. Santorini, and L. Delfs (2004). The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania.
- Kroch, A., B. Santorini, and C. Diertani (2016). The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania.
- Kroch, A., A. Taylor, and B. Santorini (2000). The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition.
- Kuteva, T., B. Heine, B. Hong, H. Long, H. Narrog, and S. Rhee (2019). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Lakoff, R. (1974). Remarks on ‘this’ and ‘that’. In *Papers from the tenth regional meeting of the Chicago Linguistic Society*, Chicago, pp. 345–356. University of Chicago.

- Lass, R. (1997). *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- Ledgeway, A. (2020). Rethinking Microvariation in Romance Demonstrative Systems. In A. Bárány, T. Biberauer, J. Douglas, and S. Vikner (Eds.), *Syntactic Architecture and its Consequences. II. Between Syntax and Morphology*, Open Generative Syntax, pp. 451–490. Berlin: Language Science Press.
- Ledgeway, A. and J. C. Smith (2016). Deixis. In *The Oxford Guide to the Romance Languages*. Oxford: Oxford University Press.
- Levinson, S. C. (2004). Deixis and pragmatics. In L. H. . G. Ward (Ed.), *The Handbook of Pragmatics*, pp. 97–121. Oxford: Blackwell.
- Lyons, J. (1975). Deixis as the source of reference. In E. Keenan (Ed.), *Formal semantics of natural language*, pp. 61–83. Cambridge: Cambridge University Press.
- Lyons, J. (1977). *Semantics*, Volume 2. Cambridge: Cambridge University Press.
- Lyons, J. (1995). *Linguistic semantics: An introduction*. Cambridge University Press.
- Marchello-Nizia, C. (1995). *L'évolution du français. Ordre des mots, démonstratifs, accent tonique*. Paris: Armand Colin.
- Marchello-Nizia, C. (2003). Se voz de ceste ne voz poéz oster, Je voz ferai celle teste copier (Ami et Amile 753) : la sphère du locuteur et la déixis en Ancien français. In A. Vanneste, P. D. Wilde, and S. Kindt (Eds.), *Memoire en temps advenir : Hommage à Th. Venckeleer*, pp. 413–427. Louvain: Peeters.

- Marchello-Nizia, C. (2005). Deixis and subjectivity: the semantics of demonstratives in Old French (9th–12th century). *Journal of Pragmatics* 37(1), 43–68.
- Meyer-Lübke, W. (1895). *Grammaire des langues romanes. Tome deuxième: Morphologie (traduction française par Auguste Doutrepon et Georges Doutrepon)*. Paris: Welter.
- Milsark, G. (1974). *Existential Sentences in English*. Ph. D. thesis, MIT.
- Mitchell, B. and F. C. Robinson (2001). *A guide to Old English*. Oxford: Blackwell.
- Næss, r., A. Margetts, and Y. Treis (2020). Introduction: Demonstratives in discourse. In Åshild Næss, A. Margetts, and Y. Treis (Eds.), *Demonstratives in discourse*, pp. 1–20. Berlin: Language Science Press.
- Patil, U., P. Bosch, and S. Hinterwimmer (2020). Constraints on German diese demonstratives: language formality and subject-avoidance. *Glossa: a journal of general linguistics* 5(1), 1–14.
- Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.
- Pedersen, H. (1905). *Les pronoms démonstratifs de l’ancien arménien: avec un appendice sur les alternances vocaliques indo-européennes*, Volume 6. B. Lunos bogtr.
- Perseus. Perseus Digital Library.
- Piotrowska, A. and D. M. Skrzypek (2021). Familiar vs. unique in a diachronic perspective. Case study of the rise of the definite article in North Germanic. *Glossa: a journal of general linguistics* 6(1), 1–21.

- Piwek, P., R.-J. Beun, and A. Cremers (2008). “proximal” and “distal” in language and cognition: evidence from deictic demonstratives in dutch. *Journal of pragmatics* 40(4), 694–718.
- Potts, C. and F. Schwarz (2010). Affective “this”. *Linguistic Issues in Language Technology* 3.
- Raw, B. C. (1997). *Trinity and incarnation in Anglo-Saxon art and thought*, Volume 21. Cambridge: Cambridge University Press.
- Rocca, R. and M. Wallentin (2020). Demonstrative reference and semantic space: a large-scale demonstrative choice task study. *Frontiers in Psychology* 11, 629.
- Rommetveit, R. (1968). *Words, Meaning, and Messages: Theory and Experiments in Psycholinguistics*. New York: Academic Press.
- Samarin, W. J. (1967). *A Grammar of Sango*. The Hague: Mouton.
- Sankoff, G. and P. Brown (1976). The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 3, 631–666.
- Schwarz, F. (2015). Topics & situations: Comments on Hinterwimmer. In *NELS 40: Semantics Workshop on Pronouns*, pp. 109–126. University of Massachusetts, Graduate Linguistic Student Association Amherst.
- Schwarz, F. (2019). Weak vs. strong definite articles: Meaning and form across languages. In A. Aguilar-Guevara, J. Pozas Loyo, and V. Vázquez-Rojas Maldonado (Eds.), *Definiteness across languages*, Number 25 in Studies in Diversity Linguistics. Language Science Press.

- Schwarz, F. (2023). Exploring the contrast between weak and strong definite articles experimentally. Talk given at New York Philosophy of Language Workshop NYU.
- Simonenko, A. (2014). *Grammatical Ingredients of Definiteness*. Ph. D. thesis, McGill University.
- Simonenko, A. and A. Carlier (2020). Between demonstrative and definite: A grammar competition model of the evolution of french *l*-determiners. *Canadian Journal of Linguistics* 65(3), 393–437.
- Skrzypek, D. (2012). *Grammaticalization of (in)definiteness in Swedish*. Wydawnictwo Naukowe UAM.
- Sommerer, L. (2011). *Old English se from Demonstrative to Article. A usage-based study of nominal determination and category emergence*. Ph. D. thesis, Universität Wien.
- Sornicola, R. (2011). Per la storia dei dimostrativi romanzi: i tipi neutri [tso],[so],[ço],[tfo] e la diacronia dei dimostrativi latini. *Zeitschrift für romanische Philologie* 127, 1–80.
- Stevens, J. (2008). Semantic Change and the Old English Demonstrative. Master’s thesis, The Ohio State University.
- Struik, T. and A. Van Kemenade (2022). Information structure and OV word order in Old and Middle English: A phase-based approach. *The Journal of Comparative Germanic Linguistics* 25(1), 79–114.
- Swanton, M. J. (1996). *The Anglo-Saxon Chronicle*. London: J. M. Dent.

- Taylor, A., A. Warner, S. Pintzuk, and F. Beths (2003). The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). Department of Linguistics, University of York. Oxford Text Archive, first edition.
- Traugott, E. C. (1980). Meaning-change in the development of grammatical markers. *Language sciences* 2(1), 44–61.
- Van Gelderen, E. (2007). The definiteness cycle in Germanic. *Journal of Germanic Linguistics* 19(4), 275–308.
- Van Gelderen, E. (2024). *The linguistic cycle: Economy and renewal in historical linguistics*. Routledge.
- Vänänen, V. (1981). *Introduction au latin vulgaire (troisième édition revue et augmentée)*. Paris: Klincksieck.
- Vanelli, L. (1997). Personal pronouns and demonstratives. In M. Maiden and M. M. Parry (Eds.), *The dialects of Italy*, pp. 106–113. London: Routledge.
- Wackernagel, J. (2009 (1918)). *Jacob Wackernagel, Lectures on Syntax: With Special Reference to Greek, Latin, and Germanic [translated by David Langslow]*. Oxford: Oxford University Press.
- Weinrich, H. (1993). *Textgrammatik der deutschen Sprache*. Mannheim: Du-den.
- Wiltschko, M. (2012). What does it take to host a (restrictive) relative clause? *Working Papers of the Linguistics Circle of the University of Victoria* 21(2), 100–145.
- Zulaica-Hernández, I. and J. Gutiérrez-Rexach (2011). On the Information Status of Antecedents: Referring Expressions Compared. In *Proceedings of*

the Workshop Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena, Göttingen, Germany, pp. 23–25.

Notes

¹Co-funded by the European Union (ERC, CAUSALITY, 101042427). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

²I am very grateful to Edith Aldridge, an anonymous reviewer, and Ion Giurgea for many helpful suggestions.

³Greenberg reserves the term deictic to Bühler's visual deictic, distinguishing it from anaphora.

⁴As a reviewer rightly notices, this is part of a more general evolution of case system in historical English. There is currently however no quantitative study of the evolution of nominal case marking with which the determiner data could be compared.

⁵“Definiteness” is used here in the sense of the property of a noun phrase to require, for its felicitous interpretation, a common ground that entails the existence of a unique individual with the property denoted by the (modified) NP in an *identifiable* situation. This term is meant to cover indiscriminately more specific conditions of use of definite articles, adnominal demonstratives, as well as adnominal possessive pronouns.

⁶Ledgeway 2020:453 puts Moldovan and Southern Daco-Romance together as having *aista/ăla* forms. According to Ion Giurgea (p.c.) (see also Giurgea (2012)), the two varieties feature different paradigms, as reflected in the mod-

ified table (Moldovan here referring to a variety within Romania).