

Article

Beyond Mere Text Frequency: Assessing Subtle Grammaticalization by Different Quantitative Measures. A Case Study on the Dutch *Soort* Construction

Robbert De Troij ^{1,2}  and Freek Van de Velde ^{1,*} ¹ Department of linguistics, KU Leuven, 3000 Leuven, Belgium; robbert.detroij@kuleuven.be² Centre for Language Studies, Radboud University Nijmegen, 6500 HD Nijmegen, The Netherlands

* Correspondence: freek.vandeveld@kuleuven.be

Received: 8 July 2020; Accepted: 30 October 2020; Published: 9 November 2020



Abstract: Grammaticalization has proven to be an insightful approach to semantic-morphosyntactic change within and across languages. Many studies, however, rely on assessing the large, obvious differences before and after the change. When investigating burgeoning or ongoing grammaticalization processes, it is notably harder to objectively measure the degree of grammaticalization. One approach is to gauge changes in the well-known ‘parameters’ of Lehmann, Hopper, and Himmelmann, but this approach is often qualitatively oriented. Quantitative studies mainly rely either on token frequency of a construction, assuming that grammaticalization is accompanied by a frequency increase, or by tracing the development of two competing constructions, looking at the proportion of their respective token frequencies. In this article, we argue for a wider range of quantitative measures, beyond token frequency, as dependent variables. We will show that these measures can jointly point to subtle ongoing grammaticalization. As a case study, we will focus on Dutch binominals with *soort* ‘sort’, a core member of the much-discussed sort-kind-type (SKT) construction in the languages of Europe. Based on a large dataset of over 14,000 instances from the period between 1850 and 1999, we investigate quantitatively measurable changes in the construction’s surface behavior (i.e., the gradual loss of the relator *van* ‘of’ and increasing restrictions on premodification and pluralization, pointing to a process of ‘decategorialization’). In addition, we will also use Gries’s deviation of proportions (DP) to gauge the dispersion of *soort*, a valuable but under-used metric in quantitative studies of grammaticalization.

Keywords: grammaticalization; SKT construction; binominals; dispersion

1. A Quantitative Approach to Grammaticalization

Grammaticalization is the process whereby new grammar is created from erstwhile lexical material in specific constructions. A good example is the adverbial suffix *-ly* in English, which derives from a noun meaning ‘shape’ or ‘body’ (Proto-Germanic **lika-*). In the vast literature on the topic that has accumulated over the years, the concept has at times been delineated more narrowly, e.g., by setting it off against lexicalization (e.g., Wischer 2000; Brinton and Traugott 2005), and in other cases has been conceived more broadly as not only capturing the process from lexical to grammatical, but also encompassing the shift from grammatical to even-more-grammatical. Recent approaches have taken a constructional perspective and have integrated grammaticalization in a process of constructionalization (Traugott and Trousdale 2013). This is not the place to review the various conceptions of grammaticalization (for good introductions and overviews, see Hopper and Traugott 2003 and Narrog and Heine 2011). Suffice it to say that the origin of the process need not necessarily be fully lexical, and the end product need not necessarily be a bound

morpheme. Take, for instance, the grammaticalization of Old English *willan* ‘wish’ to Present-day English *will* as a future-marking auxiliary. It is often considered as a case of grammaticalization (see, e.g., [Krug 2011](#), p. 554), but it is a less clear example than the adverbializing suffix *-ly*, from **lika-*. As a marker of volition, *willan* already had burgeoning auxiliary status, as it could be combined with an infinitive in Old English, as in (1). To argue that the ‘new’ function of *will* is more grammaticalized, a number of changes in its morphosyntactic behavior can be pointed out, as well as changes on the semantic level. The morphosyntactic changes have to do with what [Hopper \(1991\)](#) has called *decategorialization*: the future marker *will* gradually loses the characteristics that betray its lexical origin as a run-of-the-mill verb, such as, e.g., *do*-support.

- (1) sé ðe wyle sōð specan (*Beowulf*, v. 2864)
 he that will true speak
 ‘he who wants to speak the truth’

With cases of grammaticalization like Proto-Germanic **lika-* to *-ly*, the shift is sufficiently clear and convincing: a free morpheme with lexical meaning transforms into a bound morpheme with grammatical meaning. In the case of the transformation of *willan* to *will*, however, things are less clear. Both are free morphemes, and they both hover between fully lexical and fully grammatical. How are we to assess that one is more grammatical(ized) than the other?

There seems to be a broad consensus in the literature that grammaticalization should not be considered as a binary feature, where a construction is either ‘grammaticalized’ or ‘not grammaticalized’, but rather as a continuum. Lexical constructions slowly shift into grammatical constructions, often imperceptibly (see [De Smet 2009, 2012](#); [Van de Velde and van der Horst 2013](#); [Van der Horst and Van de Velde 2016](#), on the sneaky nature of language change), and once these constructions are in the grammar, they gradually worm their way further into the grammatical core. With regard to auxiliaries, for instance, [Heine \(1993\)](#) observes that there is no clear boundary between lexical and auxiliary verbs. [Bolinger \(1980, p. 297\)](#) noted that “[t]he moment a verb is given an infinitival complement, that verb starts down the road of auxiliariness. It may make no more than a start or travel all the way”.

It would be helpful if we had some criteria to measure the degree of grammaticalization. In principle, the diagnostics in [Lehmann \(2002\)](#) (attrition, paradigmization, obligatorification, condensation, coalescence, fixation), [Hopper \(1991\)](#) (layering, divergence, specialization, persistence, decategorialization), and [Himmelman \(2004\)](#) (host-class expansion, syntactic expansion, semantic-pragmatic expansion) are amenable to quantify the degree of grammaticalization, though in practice, this is rarely done. Exceptions are the basic operationalizations in [Bybee et al. \(1994\)](#), the corpus-based approaches in [Cheshire \(2007\)](#) and [Brems \(2011\)](#), the more radically data-driven approach in [Correia Saavedra \(2019\)](#), and the rapprochement between variationist linguistics and grammaticalization theory in, among others, [Sankoff \(1990\)](#), [Schwenter \(1994\)](#), [Poplack and Tagliamonte \(1999\)](#), [Torres-Cacoullos \(1999\)](#), [Poplack \(2011\)](#), [Wolk et al. \(2013\)](#), [Rosemeyer and Grossman \(2017\)](#), [Denis and Tagliamonte \(2018\)](#), and [Petré and Van de Velde \(2018\)](#). Sankoff, Schwenter, Poplack and Tagliamonte, Cheshire, and Denis and Tagliamonte do not provide real-time diachronic data, however, and Poplack runs different regression analyses for different periods, and, like most other variationists studies (see [Wolk et al. 2013](#), for an exemplary study), uses semantic features as independent variables, restricting the formal (dependent) variable to the relative frequency of competing variants.

Indeed, when approached from a quantitative perspective, the formal side of grammaticalization is often measured by the simple proxy of token frequency. The use of token frequency comes basically in two guises: either the rise of construction is traced through normalized frequency (the number of instances per fixed number of tokens, most often 10,000 or 1,000,000), or two competing constructions are set off against each other, and the shifting proportion of their respective token frequencies is seen as a new form encroaching on an older one, which is gradually ousted. The latter approach takes its inspiration from variationist linguistics (see [Pintzuk 2003](#)), and its preferred method is logistic regression. The underlying idea in both approaches is that increased grammatical status goes hand in

hand with increased frequency (pace some cases where rapid or instant grammaticalization occurs on the basis of analogy; see Hoffmann 2004 and Aaron 2016). While informative, this is a brutally restricted view on how to approach language change quantitatively (see Szmrecsanyi 2016 and Hilpert 2017, for well-taken considerations). The wrong way to go about this is to abandon the quantitative approach altogether. Instead, we are better off looking beyond mere tallying how often a construction occurs per million words or how often a construction occurs vis-à-vis a competing construction. This will allow us to get a better view on the different changes that transpire in the process of grammaticalization. We do not deny the pertinence of the language-internal and language-external variables that variationist studies have included in their studies, of course, but we put forward a more encompassing view on grammaticalization by looking at other formal correlates of grammaticalization. The formal behavioral and coding properties (see Haspelmath 2010, for this terminological pair) are then treated as *dependent* variables in their own right. This approach diverges from adding semantic-pragmatic (e.g., animacy, topicality), contextual (main clauses vs. subordinate clauses), and language-external variables (age, gender) as *independent* variables in a multivariate approach. The advantage of our approach is that we can pick up on subtle trends, i.e., those grammaticalization processes in which token frequency is not visibly on the rise in the time window under investigation. In this proof-of-concept article, we will look at a particular change in which we also have traditional token frequency measures. In that way, we can compare the rise in the token frequency with other measures. Our quantitative approach to grammaticalization focuses on Dutch binominals with *soort* ‘sort’. In contrast to earlier studies on the development of this Dutch construction, we will adduce quantitative diachronic evidence on the changing syntax, using frequency measures, including ‘dispersion’, a valuable metric in quantitative approaches to grammaticalization, but one not very commonly used at present.

2. The Construction at Hand

As in many other European languages, Dutch ‘binominal’ noun phrases (NPs) with *soort* ‘sort’ as first noun (N₁) are remarkably prone to grammaticalization and subjectification (see, e.g., contributions in Brems et al. 2016). The examples in (2)–(6) illustrate various uses of *soort* that coexist in Present-day Dutch, representing different stages of grammaticalization.¹

- (2) Ze eet maar twee soorten brood. (CGN, fv400275)
 she eats only two sorts bread
 ‘She only eats two types of bread.’
- (3) Hij wordt gewoon met een smoes naar de Cambrinus gehaald en dan
 he is just with an excuse to the Cambrinus lured and then
 is d’r een soort van ja afscheidsfeestje. (CGN, fn000261)
 is there a sort of yes farewell_party
 ‘He’s just lured to the Cambrinus with an excuse and then there’s a kind of farewell party.’
- (4) Dus alle tegeltjes d’raf en de pot d’ruit en zo en dat soort
 so all tiles off_it and the pot out_it and so and that sort
 dingen allemaal [...]. (CGN, fn000259)
 things all
 ‘So remove all the tiles and the pot and so on and all that kind of stuff.’
- (5) En daarna is ‘t gewoon soort van onmogelijk geworden. (CGN, fn000435)
 and after_that is it just sort of impossible become
 ‘And after that it sort of just became impossible.’

¹ The examples are taken from the Spoken Dutch Corpus (*Corpus Gesproken Nederlands* (CGN), see Oostdijk 2002).

- (6) A: Is 't ijs of niet? (CGN, fn000438)
 is it ice or not
 B: Ja soort van.
 yes sort of
 A: 'Is it ice or not?' B: 'Yes, sort of.'

In (2), *soort* is clearly used as the syntactic head, and the second noun (N_2 , *brood*) as its dependent. The construction still has its fully lexical meaning of 'subclass': it is used to profile two members of the superordinate category 'bread' that the person in question may eat. This is not the case in (3), where the N_2 (*afscheidsestje*) is the head, and the chunk (*een*) *soort* (*van*) is a qualifying modifier. Here, *soort* does not refer to some class of entities, but it is used as a hedge (downtoner, approximator), signaling that the entity the construction refers to is a less (proto)typical member of the category designated by the second noun (cf. Schermer-Vermeer 2008, pp. 20–21).² In (4), the binominal with *soort* is used as a so-called 'general extender' (GE) (Overstreet 1999, 2014; also see Van der Wouden 2014 for a discussion of GEs in Dutch), which is typically introduced by a conjunction (*en*), features a non-specific N_2 (*dingen*), and occurs usually at the end of a list. They function as "indexical expressions encoding explicit reference to further [potential members] that share with the explicit elements a common context-dependent property P" (Mauri and Sansò 2018, p. 13).

In addition to these three 'binominal' uses, *soort* can also occur outside the NP. In (5), for instance, the sequence *soort van* does not modify a noun, but an adjective (*onmogelijk*). Finally, in (6), *soort van* does not modify another constituent, but it occurs 'independently' at the clausal level, hedging the entire proposition uttered by speaker A. It is not clear whether these two uses of *soort* instantiate autonomous developments in Dutch or whether they are calques from English, where similar constructions with *sort/kind of* are (much) more prevalent (cf. e.g., Aijmer 2002, pp. 173–209). The 'independent use' is colloquial and, as such, is unlikely to crop up in corpora of written language, which makes its rise hard to pin down in time. If it is indeed a calque from English, it ought to have risen after World War II, when Dutch underwent drastic lexical influence from English (Van den Toorn 1997, pp. 559–60; Van der Sijs 2019, pp. 204–6). Our view on the matter is further complicated by the fact that native developments can be catalyzed by kindred constructions in English (see Zenner et al. 2018). This may well have been the case with the independent use of *soort* (*van*) as well: it may be an indigenous development, but with support from English. One argument in favor of the indigenous development account is that we see similar developments in *verre van* (far-from), which can be used more comfortably in independent use in Dutch than in English (see Van Goethem et al. 2018).³ We refrain from taking a strong stance with regard to the borrowed nature of this construction.

Soort has received quite some attention in the literature on Dutch nominal syntax, because of several syntactic and semantic properties that set it off against other binominals in Dutch (Van der Lubbe 1958; Vos 1999; Schermer-Vermeer 2008; Broekhuis and den Dikken 2012, pp. 631–37).⁴ However, something that has received only short shrift is the fact that, in Dutch, the preposition linking the two nouns of the binominal appears to be optional: in addition to a prepositional schema [*soort van* N], as in (3) above, Dutch sports another variant where there is no preposition, [*soort* N], as in (2) and (4).

If binominal constructions with *soort* undergo grammaticalization, as in English, we may expect decategorialization (cf. *supra*). Over time, *soort* may be expected to lose its characteristics of a typical noun, as the binominal aggregate will experience a loss in transparency (see Ten Wolde 2018 for extensive illustrations in various binominal constructions in English). The question then arises what

² The filled pause (*ja*) before the noun *afscheidsestje* in (3) can be considered another indication of this.

³ We thank one of the anonymous reviewers for this observation.

⁴ Apart from *soort*, other (semantically similar) nouns such as *type* 'type', *genre* 'genre', and *slag* 'kind' exhibit similar behavior (Schermer-Vermeer 2008, p. 5), forming in fact a small family of constructions. However, we choose to only focus on *soort* because it is the construction's most frequent and prototypical member as well as the oldest noun to have been attested in this construction (Schermer-Vermeer 2008, p. 11; Hoeksema 2011, p. 153–54).

this means for the use of the prepositional ‘relator’ between N_1 and N_2 (see Van de Velde 2009, Chp. 3 for this term). We cannot fall back on previous studies on the English sort-kind-type (SKT) construction, as English does not allow the preposition-less combinations of N_1 and N_2 in binominal construction. Dutch does allow such combinations, but not across the board (see Van de Velde 2009, Chp. 3).

Two scenarios can be envisaged. Either the bare construction [*soort* N] is pressured into the mould of a construction with a prepositional postmodifier: [*soort van* N]. This would bring the construction in line with other postmodifiers introduced with a relator. In this scenario, the use of the preposition would increase over time. Alternatively *soort* (*van*) may develop into some kind of premodifier, and it may then lose its preposition. This process has been witnessed for *miljoen* ‘million’ in the course of its development from noun to cardinal (see Van der Horst and Van de Velde 2016, pp. 412–13). If we are dealing with grammaticalization, the latter scenario is more plausible, but our case would be stronger if we additionally have other indications of increasing grammaticalization. Ideally, we want those indication to be quantitatively measurable.

What might these indications be? One additional indication is the loss of premodification of *soort* and the possibility for it to be used in the plural form. Again, this can be considered a case of decategorialization: if *soort* grammaticalizes into a premodifier of what was originally a dependent noun, then it will lose its head status and will be less likely to be premodified or be used in the plural. Second, with increasing grammaticalization, we not only expect a rise in the relative frequency of *soort*, but also a drop in its *dispersion*. Dispersion is a technical term for how equal the distribution of the construction is in a (sub)corpus. The more grammatical an element is, the less clustered it occurs, and the lower its dispersion will be.

Before we proceed to our analysis of what happens to *soort* in its ongoing grammaticalization in Late Modern Dutch, we first look into the historical provenance of the construction.

3. Historical Context

Up to Middle Dutch, the interpretative relation between two nouns would typically be encoded formally by means of inflectional morphology, viz. case marking. Speakers would ordinarily use the partitive genitive, as in (7) and (8) (Stoett 1923, pp. 102–3; Scott 2011, p. 106).

- (7) *een dropel water-s* (Stoett 1923, p. 102)
 a drop water-GEN
 ‘a drop of water’
- (8) *een pont speck-s* (Van der Horst 2008, p. 575)
 a pound bacon-GEN
 ‘a pound of bacon’
- (9) *ene geheele sorte gud-es* ((De Vries, Matthias, and Lammert Allard te Winkel 1882–1998), s.v. *soort*)
 a whole sort good-GEN
 ‘a whole lot of good (things)’

However, as a result of the erosion of case inflection in a general drift of ‘deflection’ (Weerman and de Wit 1999; Van der Horst 2013) the genitive could not be upheld, and the interpretative link between both nouns became jeopardized (Van Loey 1970, pp. 117–18; Harbert 2007, p. 90; Van der Horst 2008, pp. 143–44, among others).

As case morphology was not productive anymore, these partitive genitives were mostly replaced with periphrastic prepositional constructions (Weerman and de Wit 1999; Hoeksema 2014) or in some cases just ‘bare’ binominal NPs, as in (10) (Scott 2014, pp. 101–2).⁵

⁵ Not all genitival morphology slipped through the cracks of history. Some remnants of the genitive’s obsolescent morphological material survive in fixed expressions or have been recruited for renewed use in other functional contexts (Hoeksema 1998; Scott 2011, 2014, pp. 159–205; Norde and Trousdale 2016, pp. 174–81).

- (10) *een ketel varve* (Middle Dutch, Weerman and de Wit 1999, p. 1159)
 a kettle paint.Ø
 ‘a kettle of paint’

Many types of N_1 display (diachronic) variation, whereby the binominal NP in which they occur can be a bare one (example (11)) or a prepositional one (example (12)), or they can even occur as the second noun of a compound (example (13))—either with a slight meaning difference or not (Joosten and Vermeire 2006; Van de Velde 2009, pp. 101–5).

- (11) *een verzameling postzegels*
 a collection stamps
 ‘a collection of stamps’
- (12) *een verzameling van postzegels*
 a collection of stamps
 ‘a collection of stamps’
- (13) *een postzegel verzameling*
 a stamp collection
 ‘a collection of stamps’

Furthermore, the (diachronic) preference for one of these construction types seems to depend strongly on the identity of N_1 itself (Hoeksema 2014).⁶ However, this diachronic variation has hitherto not been charted in great detail, neither for binominal NPs with *soort*, nor for other ones, leaving us only with cautious assumptions (but see Hoeksema 2014, for a recent exception). Given that we only take into account binominal NPs with *soort*, future research will have to delve into this variation with regard to other binominal NPs as well.

4. Data

For the data collection, we turned to a newly compiled corpus of historical Dutch that covers the period between 1837 and 1999, totaling approximately 200 million tokens in size. This *Dutch Corpus of Contemporary and Late Modern Periodicals* (Dutch C-CLAMP), comprises materials from a range of cultural and literary periodicals published in Flanders and the Netherlands (Piersoul et al. 2020). It is intended as representative of standard Dutch as used by the cultural and literary elite in the period at hand.⁷ For the present study, we restricted ourselves to the 1850–1999 issues of the periodical *De Gids*, which makes up almost half of the Dutch C-CLAMP.

We used an untagged version of the corpus, from which all instances of the noun *soort* and its morphological and orthographic variants were automatically retrieved, yielding 18,964 hits. Next, false positives and doubles were manually weeded out, as were instances inside citations from other works. Incidentally, these often manifested instances of *soort* from earlier language stages, as in (14)—from Constantijn Huygens’ *Koren-bloemen* (1658) quoted in an article from 1889—which demonstrates that formerly, *soort* and N_2 (in this case, *geraecktheit*) could appear discontinuously, suggesting a stronger syntactic independence (cf. Section 3 supra, and Van de Velde 2009).

- (14) *Jan gaet aen een blaeuw oogh, en seght m’ hem niet geraecktheit; Maer dat het (en ‘t is waer)*
een soort is van geraecktheit.
 a sort is of injury
 ‘a sort of injury’

⁶ However, as Hoeksema (2014) proves for the prepositional binominal NPs, there is also a preference on the level of the construction itself, in addition to the preference of the noun that can fill N_1 , and these preferences often show diachronic fluctuations. Thus, he shows how currently the [N *aan* N]-construction is very productive—steadily gaining ground for the last century or so: [*overvloed van* N] > [*overvloed aan* N] (Hoeksema 2014, pp. 426–27).

⁷ See Willemyns (2013, pp. 110–80) for a detailed description of the external language history in the North and South of the language area in the 19–20th century.

We also removed a number of instances that featured dialectal or colloquial language (e.g., in short stories by the Flemish naturalist author Cyriel Buysse), as well as a handful of instances where there is no variation with regard to the appearance of the preposition *van*. In (15), the coordination of the nouns *vormen* and *soorten* all but blocks the omission of *van*, as the noun *vorm* obligatorily requires a *van*-complement (cf. **zovele vormen muzikaliteit*). In (16), *N*₂ is preceded by the indefinite article *een*, which similarly renders a construction without *van* ungrammatical (cf. **een soort een bocht*).

- (15) *Er zijn zovele vormen en soorten van muzikaliteit [...].*
 there are so_many forms and sorts of musicality
 (De Gids, 1950)
 ‘There are so many forms and sorts of musicality.’
- (16) *De kust maakt daar bij Hilligermond een soort van een bocht.*
 the coast makes there by Hilligermond a sort of a turn
 (De Gids, 1881)
 ‘Near Hilligermond, the coastal line makes a sort of turn.’

After the manual clean-up, we were left with 14,469 valid instances. These were all cases of binominal uses of *soort*, i.e., classifying and premodifying uses as in (2)–(4) supra. We did not encounter NP-external uses such as (5) or (6), which, as mentioned in Section 2, are more typical of colloquial language and, as such, are shunned in the prose of the authors represented in the Dutch C-CLAMP. Table 1 gives the distribution of the instances per decade.

Table 1. Distribution of binominals with *soort* per decade.

Decade	Absolute Frequency
1850s	650
1860s	1084
1870s	1227
1880s	929
1890s	1043
1900s	1021
1910s	671
1920s	694
1930s	697
1940s	369
1950s	568
1960s	1010
1970s	1592
1980s	1383
1990s	1531
Total	14,469

5. Analysis and Results

5.1. Relator Use

In this section, we report on the diachronic evolution of prepositional relator use.⁸ First, we plotted for every decade the proportion of binominal NPs containing a preposition against the ones that do not. As the mosaic plot in Figure 1 indicates, a clear trend can be discerned from the data.⁹ The data show a significant negative trend, indicating that relator use is strongly pushed back.

⁸ For data wrangling, statistical analysis, and visualization, we used the open-source software R (version 3.5.3, R Core Team 2019), and the following packages: dplyr (Wickham et al. 2020), vcd (Meyer et al. 2017), and car (Fox and Weisberg 2011).

⁹ Mosaic plots are percentage stacked bar charts with different widths for each bar, proportional to the total number of observations.

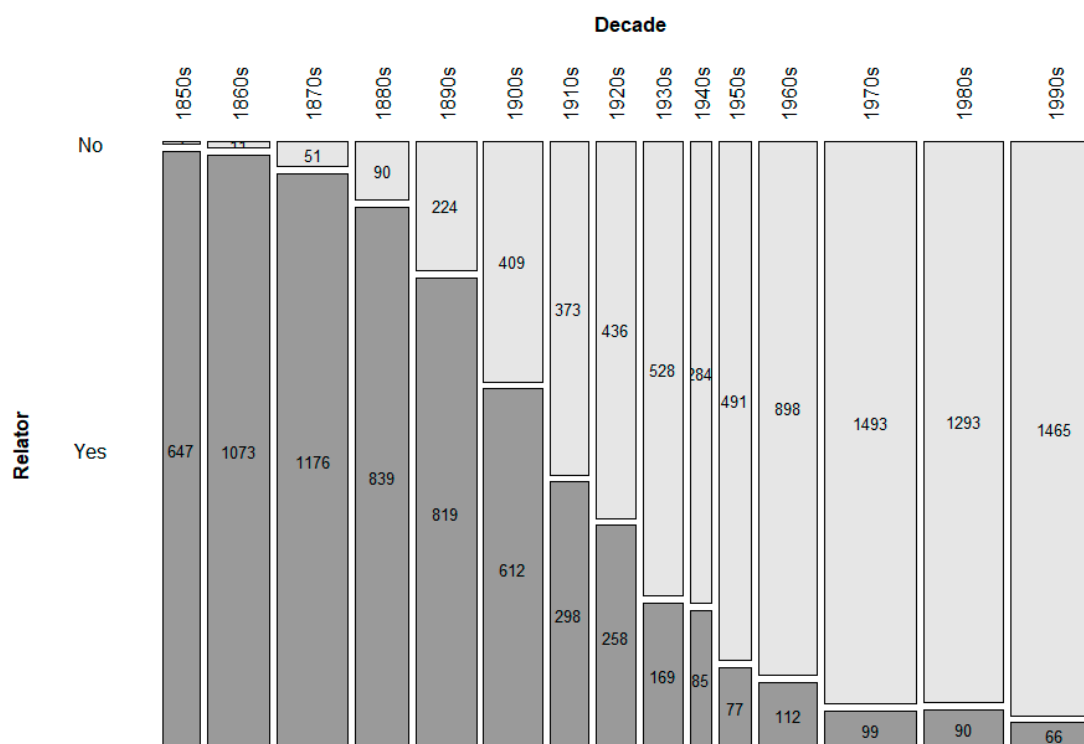


Figure 1. Diachronic trend in the proportion of relator use in binominal noun phrases (NPs) with *soort*.

The trend appears to be non-linear and conforms to the expected S-shaped trajectories that language changes typically display (see, among others, Weinreich et al. 1968, p. 113; Bailey 1973, p. 77; Kroch 1989; Labov 1994, pp. 65–67; Croft 2000, pp. 183–90; Denison 2003; Blythe and Croft 2012; Nevalainen 2015).

We fitted a logistic regression model to the data (see Baayen 2008, pp. 195–214 and Speelman 2014), predicting the presence of the relator by the year of the attestation (1850–1999). The odds of having a relator decrease through time were calculated (odds ratio: 0.95, $p < 0.001$).

The results are in line with the second scenario sketched in Section 2. The grammaticalization of *soort* involves a shift to the status of premodifier. The underlying mechanism is *reanalysis* (Harris and Campbell 1995; Hopper and Traugott 2003, p. 51): as *soort* is reanalyzed as a premodifier of N₂, with a so-called ‘hedge’ function (Lakoff 1973, p. 471) or approximative degree modifier (Traugott 2008), the preposition (*van*) originally functioning as a relator for the postmodifying N₂, becomes defunct and can be dropped. We know, from other cases of grammaticalization involving erstwhile binominal complexes, that there is typically an extensive period of variation in the use of the preposition. The next sections will elaborate on the grammaticalization of *soort* by looking at other measures.

5.2. Modification of *Soort*

As pointed out above, decategorialization in *sort*—when it shifts from head status in a fully compositional binominal construction to a premodifier with a hedging or an approximator function—will not only be visible in the loss of the prepositional relator, but potentially also in the loss of modification possibilities.

This is, however, not what we observe in the data. Figure 2 shows that there is no decreasing trend if we plot the proportion of observations in which *soort* is preceded by an adjective per decade. Indeed, a regression analysis predicting the presence of a preceding adjective on the year of attestation is nowhere near significant ($p = 0.858$). However, observe that the proportion of premodified instances of *soort* is already very low in the 1850s.

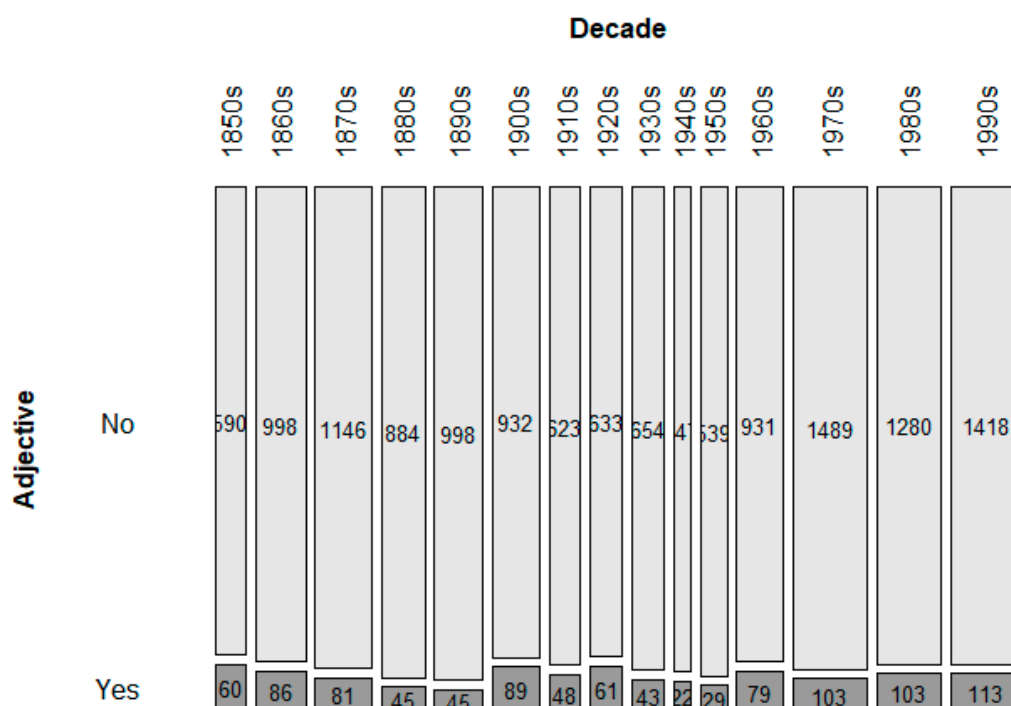


Figure 2. Adjectives preceding *soort*.

5.3. Number of *Soort*

Decategorialization of *soort* when it grammaticalizes into a hedge or approximator would lead us to expect a drop in its plural use. This is indeed what we can see in the data. Figure 3 shows a small, but significant, increase in the use of the singular. A logistic regression analysis with the year of attestation as the predictor gives an odds ratio of 1.01 ($p < 0.001$).

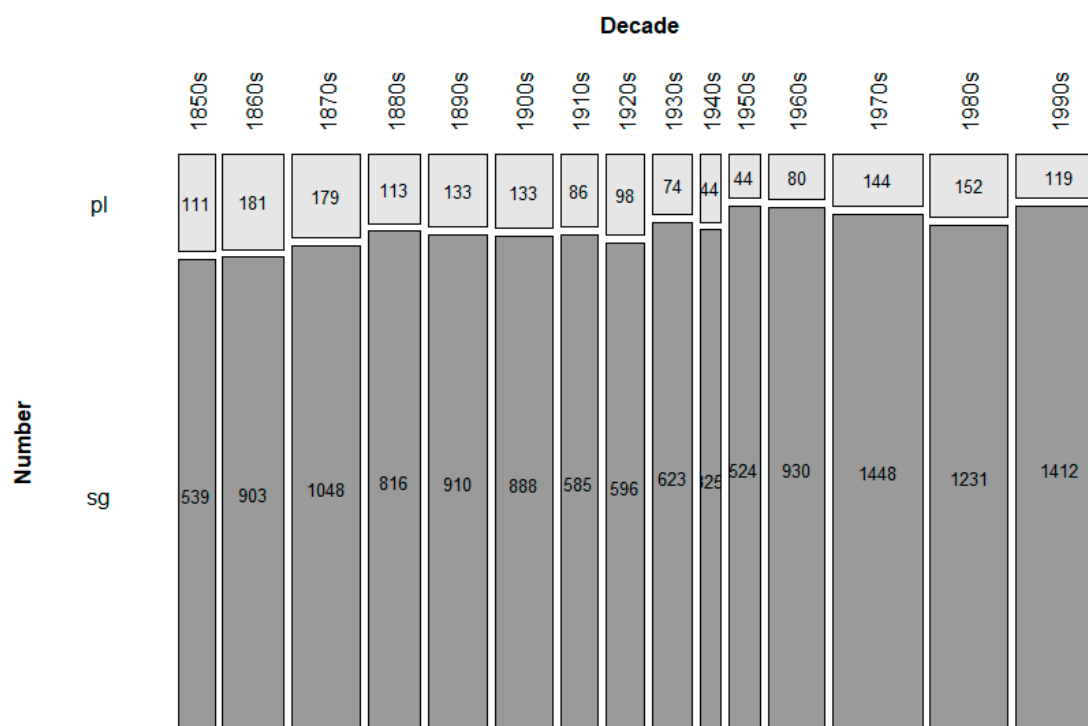


Figure 3. Singular (*soort*) vs. plural (*soorten*).

5.4. Frequency Measures

Decategorialization is a known concomitant process in grammaticalization, but, as we have just seen, the results are not fully unequivocal. The relator is increasingly dropped, and there is some indication that *soort* over time tends to become fixed in the singular, but there is no discernable trend in the loss of premodifying adjectives (although premodification of *soort* already appears to be quite rare in the 1850s). The reason that the measures are not revealing a clear signal is that the grammaticalization is subtle and spans a large time period. Detecting a trend requires us to dig deeper.

Let us turn to frequency measures to give some further support for the idea that *soort* is gradually grammaticalizing into a premodifier. As Figure 4 shows, we can see a rise of the relative frequency, though the effect seems to be mainly situated in the most recent decades. One possible, though highly tentative, language-external explanation for this sudden dramatic increase in the 1950s, after a long period of quite stable frequencies, could be the more intensive contact with the English-speaking world in the post-World-War-II period (cf. Section 2 *supra*). A linear regression predicting the relative frequency by the decade yields a slope of 1.831 ($p < 0.001$).

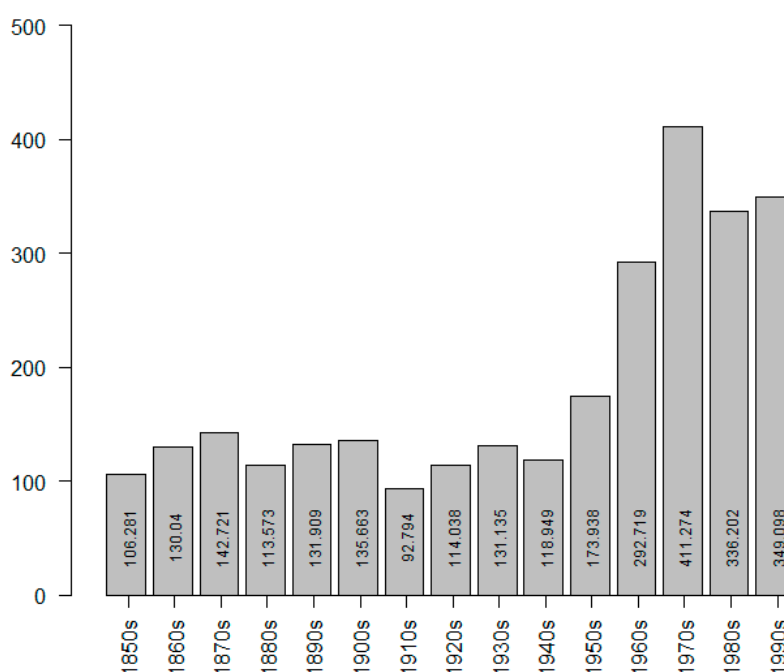


Figure 4. Relative frequency of the *soort* construction: number of observations per million tokens.

A second frequency metric is *dispersion*, a measure for how equal the distribution of a lexeme or construction is in a corpus. Constructions with a high dispersion occur in a clustered fashion, meaning they occur in bursts, with long intervals in which they are absent. Constructions with a low dispersion occur at regular intervals. Dispersion can be considered as a concomitant of increased grammaticalization (Hilpert 2017; Petré and Van de Velde 2018; Saavedra 2019; Anthonissen 2020): as elements shift from the lexical end of the cline to the grammatical end of the cline, they will be less specific, and will not be tied to a specific context. The less lexically specific an element is, the less it will be tied to a specific topic. Highly grammaticalized elements, such as auxiliaries or definite and indefinite articles, will occur inevitably in all texts, irrespective of the topic, whereas words like *nitrogen*, *Ostrogoth*, or *quivered* will tend to cluster in specific parts of the textual record. Thus, over time, the increasing promiscuity of grammaticalizing constructions will show up in a decrease in their dispersion. This is also what we expect for *soort*.

In a comparison of different DP measures, [Gries \(2008\)](#) argues that the ‘(normalized) deviation of proportions’ (DP_{Norm}) yields the best results, and this is the measure we adopt here as well. Deriving the normalized DP is accomplished by the following algorithm:

- i. Cut the corpus in n different parts. These parts may be naturally given, for instance, different text files.
- ii. Measure how large each of these parts is, in number of tokens, and express this as a percentage of the total corpus. This gives an indication of the expected frequency proportions.
- iii. Count the frequency of the lexeme or construction of interest in each of these parts and express this as a percentage of the total frequency. This gives the observed frequency proportions.
- iv. Make a pair-wise comparison between the expected and observed proportions.
- v. Sum up the proportions and divide them by 2.
- vi. Normalize by dividing DP by $1 - (1/n)$, with n = total number of parts in the corpus.

Steps (i) to (v) yield the DP, which falls in the range from 0 to 1, where values approximating 0 indicate that the element is equally distributed over the corpus parts in proportion to the size of the corpus parts, whereas values approximating 1, by contrast, are indicative of a skewed distribution, where the element clusters in specific subparts of the corpus. Step (vi) is a normalization, dampening the DP value in case there are more subparts of the corpus over which DP is measured. As n grows, the denominator $1 - (1/n)$ will grow, and the normalized DP will shrink.

We split the corpus in decades and calculated the normalized DP (DP_{Norm}) of the string *soort*. We then looked for a trend in the DP_{Norm} over time.

DP_{Norm} assumes that the different subparts of the corpus are comparable; that is, they may differ in size, as the DP takes that into account by the pairwise comparison between the observed and expected frequencies (step (iv) in the procedure above), but they should not differ according to a possibly hidden dimension, which could explain the difference in the distribution of the element under investigation apart from the effect we are interested in. To check whether the subparts are ‘commensurable’, we first calculated the DP_{Norm} of the definite article *de*. As the period under investigation is a language stage in which the article has been fully entrenched in the Dutch language (see [Van de Velde 2009](#)), we expect the DP_{Norm} of the definite article to remain stable. However, we noticed some volatility over time. To control for this slight incommensurability of the subparts, we divided the DP_{Norm} of *soort* by the DP_{Norm} of *de*. Doing so revealed a decreasing trend, as expected; see Figure 5. There is a negative correlation: linear regression on the decade indicates that this trend is significant (slope -0.011 , $p < 0.01$).

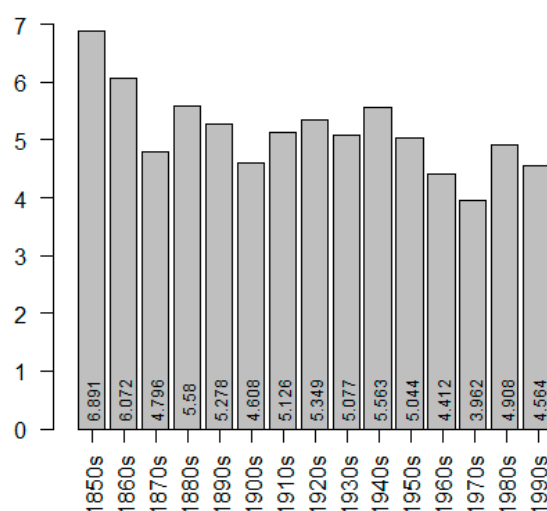


Figure 5. Normalized deviation of proportions (DP) of *soort* per decade (divided by normalized DP of *de*).

5.5. A multivariate Approach

Especially in variations studies, but increasingly also in historical linguistics, quantitative analyses often have a multivariate design. The most commonly used technique is regression analysis, with a numeric or binary outcome variable, and several independent variables, possibly in interaction (see [Van de Velde and Peter 2020](#), pp. 344–46 for an introduction, and exemplary studies cited there). In our case study on *soort*, we have several formal dependent variables and one independent variable: time. We could, however, turn this into a multivariate design by linearly regressing the numeric variable time on all the linguistic variables (see [Petré and Van de Velde 2018](#), for an earlier application). This basically constitutes a reversal of the typical approach in which time is included as an independent variable to predict the relative token frequency of a construction.

If we then do variable selection on the basis of the Akaike Information Criterion (AIC), which incurs a penalty for adding variables and puts a premium on model parsimony, we could, in principle, see which of the formal diagnostics of grammaticalization weighs in more heavily. Of course, this can only be done if the diagnostics are measured at the same level. At the level of the individual observations, we can take into account the presence of a relator *van* (Section 5.1), the presence of the preceding adjective (Section 5.2), and the number (Section 5.3). The frequency measures in Section 5.4 cannot be used, evidently, as these are not measured at the level of the individual observation.

If we run a bidirectional stepwise variable selection procedure on AIC (see [Levshina 2015](#), pp. 149–52), the best model is obtained by including all the variables as main effects, though the presence of the preceding adjective has only a limited effect on model quality, and it is not significant as a predictor itself. No interaction-effects are needed. The best model is reported in full in Table 2. It has an adjusted R^2 of 0.56, meaning that more than half of the variance is accounted for. There is no indication of heteroscedasticity (Non-constant Variance Score Test $p = 0.902$), nor of multicollinearity (all Variance Inflation Factors (VIFs) are lower than 5). There is some indication of autocorrelation (Durbin-Watson test $p < 0.001$), but with a fair number of data points, the Durbin-Watson test may be oversensitive and taking autocorrelation into account requires more complicated models (see [Van de Velde and Peter 2020](#), pp. 346–50 for discussion).

Table 2. Regression model output.

Variable	Level	Estimate	Standard Error	t Value	p Value
(Intercept)	-	1960.54	0.36	5452.61	<0.001
Presence of relator (<i>van</i>)	Absent	(base level)	-	-	-
	Present	-70.55	0.53	-133.83	<0.001
Number	Singular	(base level)	-	-	-
	Plural	4.82	0.82	5.91	<0.001
Preceding adjective	No	(base level)	-	-	-
	Yes	-1.87	1.02	-1.84	0.066

The fact that all variables are selected in the multivariate linear model and the absence of multicollinearity could be taken as indications that grammaticalization diagnostics do not measure the exact same thing. Though they may be instantiations of the same overarching phenomenon of decategorialization, they do not happen in strict synchronicity, echoing the findings in [Van der Horst and Van de Velde \(2016\)](#). Note that this does *not* bear on the discussion of the Constant Rate Effect (CRE), as advocated by [Kroch \(1989\)](#) and subsequent work (see also [Pintzuk 2003](#)). The CRE says that frequency increase may start at different points in time for different syntactic contexts

(e.g., main clauses vs. subordinate clauses), but proceeds at the same speed.¹⁰ The CRE takes into account only token frequency. It remains to be seen whether it carries over to other quantitative dependent variables.

6. Conclusions

As the prototypical member of the Dutch SKT construction, *soort* is gradually grammaticalizing into a hedge or approximator. This shift entails a rebracketing (Hopper and Traugott 2003, p. 51) from a transparent, compositional binominal construction into a premodified NP, in which the original postmodifier is now the head noun and *soort*, together with its prepositional relator *van*, is reanalyzed as a premodifier. In this process of grammaticalization, *soort* undergoes ‘decategorialization’ and loses more and more of its nominal syntax. This process spans an extensive period and happens very gradually, and the semantic nature of the shift in head status is invisible on the surface. All this makes it hard to catch the reanalysis in the act. We set out to use different quantitative measures. Rather than merely looking at token frequency, taking an observed rise to be a straightforward indication of grammaticalization, we looked at several formal diagnostics.

In our data, ranging from the 1850s to the 1990s, we can observe a clear trend of dropping *van* and a (somewhat weaker but still significant) trend of losing the possibility of it occurring in the plural, which are both indications of decategorialization, one of the defining diagnostics of grammaticalization. Another expected decategorialization trend, namely the loss of modification possibilities, was not supported by our data. The decategorialization trends turned out to be somewhat equivocal: of the three expected trends, one manifested itself clearly, the other weakly, and the third not at all. In order to support the idea that *soort* is indeed undergoing grammaticalization, we turned to two other measures that rely on frequency; namely, relative frequency and dispersion. Both trends are in line with the scenario of increasing grammaticalization: relative frequency goes up through the years and dispersion goes down.

In sum, gathering a substantial number of observations, we were able to pick up a subtle grammaticalization trend that would otherwise have been hard to detect. A possible objection (and maybe an explanation for the unequivocal results for some of the expected trends) is that we did not separate the data on the basis of semantics, more specifically whether *soort* has a classifying or premodifying function, and that our aggregate results may be obfuscated by a ‘layering’ effect (in the sense of Hopper 1991), in which the pattern in the old non-grammaticalized function continues to be used alongside the new function. However, this is counterbalanced, we feel, by the objectivity of our methods. Categorizing data according to the function they have, in a non-circular way, is a subjective enterprise.

Author Contributions: Conceptualization, R.D.T. and F.V.d.V.; methodology, R.D.T. and F.V.d.V.; investigation, R.D.T. and F.V.d.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the BOF-project ‘How predictable is language change?’ (grant number C14/18/035).

Acknowledgments: We would like to thank the National Library of The Netherlands (KB) for making available its resources for linguistic research.

Conflicts of Interest: The authors declare no conflict of interest.

¹⁰ Technically: it has a different intercept but a similar estimate (‘beta’) for the time variable in the linear part of the generalized linear model.

References

- Aaron, Jessi E. 2016. The road already traveled: Constructional analogy in lexico-syntactic change. *Studies in Language* 40: 26–62.
- Aijmer, Karin. 2002. *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins.
- Anthonissen, Lynn. 2020. Special Passives across the Lifespan: Cognitive and Social Mechanisms. Ph.D. dissertation, University of Antwerp, Antwerp, Belgium.
- Baayen, Rolf Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Bailey, Charles-James. 1973. *Variation and Linguistic Theory*. Washington: Center for Applied Linguistics.
- Blythe, Richard, and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88: 269–304.
- Bolinger, Dwight. 1980. Wanna and the gradience of auxiliaries. In *Wege zur Universalienforschung: Sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler*. Edited by Gunter Brettschneider and Christian Lehmann. Tübingen: Gunter Narr, pp. 292–99.
- Brems, Lieselotte, Bernard De Clerck, and Katrien Verveckken, eds. 2016. Binominal syntagms as loci of synchronic variation and diachronic change. *Language Sciences* 53: 99–208.
- Brems, Lieselotte. 2011. *Layering of Size and Type Noun Constructions in English*. Berlin: Mouton de Gruyter.
- Brinton, Laurel J., and Elizabeth C. Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Broekhuis, Hans, and Marcel den Dikken. 2012. *Syntax of Dutch: Nouns and Noun Phrases*. Amsterdam: Amsterdam University Press, vol. 2.
- Bybee, Joan, Revere D. Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Cheshire, Jenny. 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11: 155–93.
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. London: Longman.
- De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119: 1728–55.
- De Smet, Hendrik. 2012. The course of actualization. *Language* 88: 601–33.
- Denis, Derek, and Sali A. Tagliamonte. 2018. The changing FUTURE: Competition, specialization and reorganization in the contemporary English future temporal reference system. *English Language and Linguistics* 22: 403–30.
- Denison, David. 2003. Log(ist)ic and simplistic S-curves. In *Motives for Language Change*. Edited by Raymond Hickey. Cambridge: Cambridge University Press, pp. 54–70.
- Fox, John, and Sanford Weisberg. 2011. *An R Companion to Applied Regression*, 2nd ed. Thousand Oaks: Sage.
- Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13: 403–37.
- Harbert, Wayne. 2007. *The Germanic Languages*. Cambridge: Cambridge University Press.
- Harris, Alice, and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 2010. The behaviour-before-coding principle in syntactic change. In *Essais de Typologie et de Linguistique Générale. Mélanges Offerts à Denis Creissels*. Edited by Franck Floricic. Paris: Presses de L'École Normale Supérieure, pp. 493–506.
- Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.
- Hilpert, Martin. 2017. Frequencies in diachronic corpora and knowledge of language. In *The Changing English Language—Psycholinguistic Perspectives*. Edited by Marianne Hundt, Simone Pfenninger and Sandra Mollin. Cambridge: Cambridge University Press, pp. 49–68.
- Himmelman, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In *What Makes Grammaticalization? A Look from Its Fringes and Its Components*. Edited by Walter Bisang, Nikolaus P. Himmelman and Björn Wiemer. Berlin: Mouton de Gruyter, pp. 21–42.
- Hoeksema, Jack. 1998. Een dode categorie: De genitief. *Tabu* 28: 162–67.
- Hoeksema, Jack. 2011. Het WNT: Een waarlijk nuttige tool? *Nederlandse Taalkunde* 16: 152–59.

- Hoeksema, Jack. 2014. De opkomst van aan als verbindend element in maatnamenconstructies. In *Patroon en Argument: Een Dubbelfeestbundel Bij het Emeritaat van William Van Belle en Joop van der Horst*. Edited by Freek Van de Velde, Hans Smessaert, Frank Van Eynde and Sara Verbrugge. Leuven: Leuven University Press, pp. 421–32.
- Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data—And the importance of intuition. In *Corpus Approaches to Grammaticalization in English*. Edited by Hans Lindquist and Christian Mair. Amsterdam: John Benjamins, pp. 171–210.
- Hopper, Paul, and Elizabeth Closs Traugott. 2003. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.
- Hopper, Paul. 1991. On some principles of grammaticalization. In *Approaches to Grammaticalization. Vol. 1: Focus on Theoretical and Methodological Issues*. Edited by Elizabeth C. Traugott and Bernd Heine. Amsterdam: John Benjamins, pp. 17–35.
- Joosten, Frank, and Lea Vermeire. 2006. Collectiva en relationaliteit. *Nederlandse Taalkunde* 11: 23–58.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199–244. [CrossRef]
- Krug, Manfred. 2011. Auxiliaries and grammaticalization. In *The Oxford Handbook of Grammaticalization*. Edited by Heiko Narrog and Bernd Heine. Oxford: Oxford University Press, pp. 547–58.
- Labov, William. 1994. *Principles of Linguistic Change. Vol. 1: Internal Factors*. Oxford: Blackwell.
- Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2: 458–508.
- Lehmann, Christian. 2002. *Thoughts on Grammaticalization*, 2nd revised ed. Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt. pp. 1–171. Available online: <http://files.professorivo.webnode.pt/200000047-800fd8a61f/Thoughts%20on%20grammaticalization.pdf> (accessed on 3 November 2020).
- Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Mauri, Caterina, and Andrea Sansò. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. *Folia Linguistica Historica* 39: 1–35. [CrossRef]
- Meyer, David, Achim Zeileis, and Kurt Hornik. 2017. vcd: Visualizing Categorical Data. R Package Version 1.4-4. Available online: <https://CRAN.R-project.org/package=vcd> (accessed on 5 November 2020).
- Narrog, Heiko, and Bernd Heine, eds. 2011. *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.
- Nevalainen, Terttu. 2015. Descriptive adequacy of the S-curve model in diachronic studies of language change. In *Can We Predict Linguistic Change?* Edited by Christina Sanchez-Stockhammer and Friedrich Alexander. Helsinki: VARIENG, Available online: <http://www.helsinki.fi/varieng/journal/volumes/16/nevalainen/> (accessed on 5 November 2020).
- Norde, Muriel, and Graeme Trousdale. 2016. Exaptation from the perspective of Construction Morphology. In *Exaptation and Language Change*. Edited by Muriel Norde and Freek Van de Velde. Amsterdam: John Benjamins, pp. 163–95.
- Oostdijk, Nelleke. 2002. The design of the Spoken Dutch Corpus. In *New Frontiers of Corpus Research*. Edited by Pam Peters, Peter Collins and Adam S. Cohen. Amsterdam: Rodopi, pp. 105–12.
- Overstreet, Maryann. 1999. *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse*. Oxford: Oxford University Press.
- Overstreet, Maryann. 2014. The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics* 24: 105–29. [CrossRef]
- Petré, Peter, and Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94: 867–901. [CrossRef]
- Piersoul, Jozefien, Robbert De Troij, and Freek Van de Velde. 2020. *150 Years of Written Dutch: The Construction of the Dutch Corpus of Contemporary and Late Modern Periodicals*, Unpublished manuscript.
- Pintzuk, Susan. 2003. Variationist approaches to syntactic change. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell, pp. 509–28.
- Poplack, Shana, and Sali A. Tagliamonte. 1999. The grammaticalization of *going to* in (African American) English. *Language Variation and Change* 11: 315–42.
- Poplack, Shana. 2011. A variationist perspective on grammaticalization. In *Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Blackwell, pp. 209–24.

- Rosemeyer, Malte, and Eitan Grossman. 2017. The road to auxiliarity revisited: The grammaticalization of FINISH anteriors in Spanish. *Diachronica* 34: 516–58. [CrossRef]
- Saavedra, David Correia. 2019. Measurements of Grammaticalization: Developing a Quantitative Index for the Study of Grammatical Change. Ph.D. dissertation, Université de Neuchâtel, Neuchâtel, Switzerland. University of Antwerp, Antwerp, Belgium.
- Sankoff, Gillian. 1990. The grammaticalization of tense and aspect in Tok Pisin and Sranan. *Language Variation and Change* 2: 295–312. [CrossRef]
- Schermer-Vermeer, Ina. 2008. De Soort-constructie: Een nieuw patroon in het Nederlands. *Nederlandse Taalkunde* 13: 2–33.
- Schwenter, Scott A. 1994. The grammaticalization of an anterior in progress: Evidence of peninsular Spanish dialect. *Studies in Language* 18: 71–111. [CrossRef]
- Scott, Alan. 2011. The position of the genitive in Present-day Dutch. *Word Structure* 4: 104–35. [CrossRef]
- Scott, Alan. 2014. *The Genitive Case in Dutch and German: A Study of Morphosyntactic Change in Codified Languages*. Leiden: Brill.
- Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Edited by Dylan Glynn and Justyna Robinson. Amsterdam: John Benjamins, pp. 487–533.
- Stoett, Frederik August. 1923. *Middelnederlandsche Spraakkunst: Syntaxis*. Den Haag: Martinus Nijhoff.
- Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12: 153–71. [CrossRef]
- Ten Wolde, Elnora G. 2018. Diversity and Diversification among of-Binominal Noun Phrases: The Case of the Evaluative Binominal Noun Phrase Family. Ph.D. dissertation, University of Vienna, Vienna, Austria.
- Torres-Cacoullous, Rena. 1999. Variation and grammaticalization in progressives: Spanish-ndo. *Studies in Language* 23: 25–59.
- Traugott, Elizabeth C. 2008. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*. Edited by Regine Eckhardt, Gerhard Jäger and Tonjes Veenstra. Berlin: Mouton de Gruyter, pp. 219–50.
- Traugott, Elizabeth C., and Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Van de Velde, Freek, and Johannes M. van der Horst. 2013. Homoplasmy in diachronic grammar. *Language Sciences* 36: 66–77. [CrossRef]
- Van de Velde, Freek, and Petrú Peter. 2020. Historical linguistics. In *The Routledge Handbook of English Language and Digital Humanities*. Edited by Svenja Adolphs and Dawn Knight. London: Routledge, pp. 328–59.
- Van de Velde, Freek. 2009. *De Nominale Constituent: Structuur en Geschiedenis*. Leuven: Leuven University Press.
- Van den Toorn, Maarten C. 1997. Nieuwnederlands (circa 1920–nu). In *Geschiedenis van de Nederlandse Taal*. Edited by Maarten C. van den Toorn, Willy J. J. Pijnenburg, J. Arjan van Leuvensteijn and Johannes M. van der Horst. Amsterdam: Amsterdam University Press, pp. 479–562.
- Van der Horst, Johannes M. 2008. *Geschiedenis van de Nederlandse Syntaxis*. Leuven: Leuven University Press.
- Van der Horst, Johannes M. 2013. *Taal op Drift: Lange-Termijnontwikkelingen in Taal en Samenleving*. Amsterdam: Meulenhoff.
- Van der Horst, Johannes M., and Freek Van de Velde. 2016. Miljoen. *Leuvense Bijdragen—Leuven Contributions in Linguistics and Philology* 100: 410–24.
- Van der Lubbe, Henricus. 1958. *Woordvolgorde in Het Nederlands: Een Synchrone Structurele Beschouwing*. Assen: Van Gorcum.
- Van der Sijs, Noline. 2019. *15 Eeuwen Nederlandse Taal*. Gorredijk: Sterck, and De Vreese.
- Van der Wouden, Ton. 2014. Een artikel over of zo, en zo. In *Black Book: A Festschrift for Frans Zwarts*. Edited by Jack Hoeksema and Dicky Gilbers. Groningen: University of Groningen, Available online: <http://www.let.rug.nl/hoeksema/festschrift.html> (accessed on 2 September 2020).
- Van Goethem, Kristel, Gudrun Vanderbauwhede, and Hendrik De Smet. 2018. The emergence of a new adverbial downtoner: Constructional change and constructionalization of Dutch [*ver van X*] and [*verre van X*] ‘far from X’. In *Category Change from a Constructional Perspective*. Edited by Kristel Van Goethem, Muriel Norde, Evie Coussé and Gudrun Vanderbauwhede. Amsterdam: John Benjamins, pp. 179–206.

- Van Loey, Adolphe. 1970. *Schönfeld's Historische Grammatica van Het Nederlands: Klankleer, Vormleer en Woordvorming*, 8th ed. Zutphen: Thieme.
- Vos, Riet. 1999. A Grammar of Partitive Constructions. Ph.D. dissertation, University of Tilburg, Tilburg, The Netherlands.
- Weerman, Fred, and Petra de Wit. 1999. The decline of the genitive in Dutch. *Linguistics* 37: 1155–92. [CrossRef]
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*. Edited by Winfred Lehmann and Yakov Malkiel. Austin: University of Texas Press, pp. 95–188.
- Wickham, Hadley, François Romain, Henry Lionel, and Kirill Müller. 2020. dplyr: A Grammar of Data Manipulation. R Package Version 0.8.4. Available online: <https://CRAN.R-project.org/package=dplyr> (accessed on 5 November 2020).
- Willemyns, Roland. 2013. *Dutch: Biography of a Language*. Oxford: Oxford University Press.
- Wischer, Ilse. 2000. Grammaticalization versus lexicalization: “Methinks” there is some confusion. In *Pathways of Change: Grammaticalization in English*. Edited by Olga Fischer, Anette Rosenbach and Dieter Stein. Amsterdam: John Benjamins, pp. 355–70.
- De Vries, Matthias, and Lammert Allard te Winkel. 1882–1998. *Woordenboek der Nederlandsche Taal*. The Hague: Martinus Nijhoff.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach, and Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructural variation and change. *Diachronica* 30: 382–419. [CrossRef]
- Zenner, Eline, Kris Heylen, and Freek Van de Velde. 2018. Most borrowable construction ever! A large-scale approach to contact-induced pragmatic change. *Journal of Pragmatics* 133: 134–49. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).