THE SPARK OR THE FUEL? ON THE ROLE OF AMBIGUITY IN LANGUAGE CHANGE

HENDRIK DE SMET
KU LEUVEN

MARIE-ANNE MARKEY
KU LEUVEN

ABSTRACT In many theories of language change, ambiguity is put forward as one of the main causes of change. It is thought to be the trigger for syntactic reanalysis, and it is also assumed to cause semantic change. This paper reconsiders the role of ambiguity in change, arguing that its role in change is more complicated. To show this, the role of ambiguity is investigated in the history of English *over*, as it shifted from preposition to adnumeral marker (as in *we have been discussing trivial matters for over three hours*). Although syntactic ambiguity between adnumeral and prepositional readings is common, ambiguous uses cannot be shown to significantly predate unambiguously new uses. At the same time, once adnumeral *over* began to appear in the historical record, it spread more readily to ambiguous contexts. The findings show that syntactic change does not necessarily depend on ambiguous contexts as a trigger of innovation, but the same contexts can facilitate the subsequent spread of an innovation. These conclusions fit a model of change that depends more heavily on analogy as its driving force. It is suggested that similar effects may be at play in semantic change. Finally, it is clear that the mere observation of ambiguity in the historical record is not conclusive evidence of a reanalysis-through-ambiguity type of innovation.

1 INTRODUCTION

In the who-done-it of language change, ambiguity is one of the usual suspects. Ambiguity is commonly seen as the trigger for syntactic reanalysis (Langacker 1977, Timberlake 1977, Anttila 1989, Harris & Campbell 1995, Campbell 1998, Harris 2003) and, more generally, it has been suggested that "systemic ambiguity is perhaps the major basis for syntactic change" (Hock 1986: 350). Ambiguity is also widely cited as a trigger for semantic change. Heine traces semantic change to "bridging contexts" (2002: 84), where a new

meaning is inferred even as the old source meaning "cannot be ruled out" (2002: 84). Similar views have been expressed by Croft (2000: 118), Evans & Wilkins (2000: 550), Diewald (2006: 3), and many others. As Eckardt elegantly puts it: "Turning points in language change are often defined by utterances [...] of a double-faced nature" (Eckardt 2011: 389).

However, the usual suspect cannot always be found guilty. It has been argued that ambiguity may sometimes be the result rather than the cause of change (Fischer 2007, De Smet 2009, 2013). Probably on similar grounds, Traugott & Dasher warn that "projection from present-day examples of ambiguity to potential loci for change in the past is a useful first-pass methodology, but it cannot lead to definitive reconstruction of mechanisms" (2002: 131). It has further become increasingly clear that ambiguity is not a unitary concept (Winter-Froemel 2014, Denison 2017). And, most importantly, it is by no means the case that ambiguity is involved in *all* instances of change (Traugott & Trousdale 2013: 199). All this indicates that the role of ambiguity in language change deserves further careful study.

An important question, in that respect, is how to go about evaluating the role of ambiguity in change. Winter-Froemel (this vol.) offers a detailed discussion of a number of different changes plausibly triggered by ambiguity in bridging contexts. For each change, she distinguishes between old 'conventional uses', 'bridging uses' and 'new uses' (in addition to 'other uses', which are not directly implicated in the change). She then goes on to show that bridging uses appear earlier than new uses and that, just before or at the time new uses first appear, bridging uses take up a sizeable share of all uses. In light of this, Winter-Froemel suggests that, for high-frequency items, a threshold of 50% 'bridging uses exposure' should be reached for reanalysis to be likely. While causality remains, strictly speaking, impossible to prove, the specifics of the changes discussed by Winter-Froemel conform well to a scenario where ambiguity is the trigger of innovation – or, as we will have it here, the spark that ignites change.

In our own contribution, however, we offer a counterpoint to the changes discussed by Winter-Froemel. In what follows, we focus on ambiguity in one particular change, the emergence of English adnumeral *over*, which instantiates a shift from preposition, as in (1a), to adnumeral marker, as in (1b).

(1) a. *'It's not a school any more,' he said **over** his shoulder.* (1991, BNC)

b. *A massive, grey stone Victorian building, it housed **over** 1600 inmates, twice its allotted amount.* (1992, BNC)

Of course we recognize that what can be gleaned from the history of *over* does not automatically translate to every other putative case of change through ambiguity. Nevertheless, the details of this change prove worth considering. We will argue that in this change structural ambiguity is implicated, but not in the way ambiguity is usually believed to be implicated. That is, ambiguity in this case is not the initial cause of innovation but works as a factor promoting an innovation's subsequent spread. To the fire of language change, then, ambiguity here acts not so much as the spark that starts it, but as the fuel that keeps it going.

In a number of ways, the implications of this bear more generally on our understanding of syntactic and (arguably) semantic reanalysis. On the one hand, like Traugott & Dasher (2002), we want to caution that even widespread ambiguity in the historical record is not in itself evidence that ambiguity is necessarily the initial cause of change. On the other, if in the case of *over* it is not ambiguity that is the trigger of change, we have to confront the twin questions of what else can trigger change, and how else change can be impacted by ambiguity. In what follows, we will first describe the change that led to the emergence of adnumeral *over* (2), then use extensive corpus data to document the role of ambiguity (3, 4), and discuss in greater detail the implications of our findings, offering an alternative model of change (5) and considering the further methodological and theoretical consequences (6). The argument is finally summarized in the conclusion (7).

2   FROM PREPOSITION TO ADNUMERAL

The change from preposition to adnumeral marker has been a common type of development in European languages (Plank 2004). In English, the development can, for instance, be seen in *about*, *above*, *around*, *below*, *over* and *under*. Traditional syntax does a good job describing the start and end points of this type of change, which can be thought of as a dependency shift and concomitant rebracketing. In (2), *over the warm chocolate* is a prepositional phrase (PP), doing what a PP is expected to do – in this case, function as a place adverbial. Within the PP, *over* can be assumed to be the head, being the element that determines the syntactic profile of the whole complex unit. Prepositions typically take noun phrases (NPs) as complement, here dutifully supplied by *the warm chocolate*.

(2)   *With a fork draw lines **over** the warm chocolate to resemble bark.* (s.d., BNC)

In contrast, in (3), *over ninety states* does not do what a PP normally does. It functions as subject, making it unlikely that it is a PP with *over* as its head.[1] As the finite verb *have* agrees with the plural *states*, the most plausible analysis treats *over ninety states* as an NP, with the plural noun *states* as its head. It follows that *over* modifies *ninety*, which is itself a modifier to the head *states*. Specifically, combined with a numeral, *over* profiles a region on a numerical scale above the reference value provided by the numeral.

(3)   ***Over*** *ninety states have accepted the obligation not to acquire nuclear arms under the 1968 Non-Proliferation Treaty* (1986, BNC)

In brief, a preposition heading a prepositional phrase turns into a modifier of a numeral within a noun phrase. In bracketed notation, [*over* [*the warm chocolate*]$_{NP}$]$_{PP}$ in (2) develops into [[*over ninety*]$_{NUM}$ *states*]$_{NP}$ in (3). Other consequences, and thus further evidence, of the syntactic change include the fact that *over*, as an adnumeral marker, can follow any preposition, as in (4a), can function as direct object, as in (4b), and can even be preceded by a determiner, as in (4c).

(4)   a. *The pilot selected for the mission was an experienced lieutenant-colonel with **over** 7,000 flying hours on the F-15.* (1989, BNC)

b. *They calculate they must have tried **over** seventy different shades of paint before finding the correct 'cramoisy'* (1991, BNC)

c. *The largest of the **over** 600 dwellings in the park is the Cliff Palace.* (s.d., iWeb)

In present-day usage, then, *over*, can be a preposition or an adnumeral. Importantly for present purposes, between those two uses, ambiguity is common. Ambiguity arises whenever a phrase introduced by *over* can be analysed both as PP and as NP. Typically, that depends on the function it realizes in clausal syntax. An example is given in (5).

(5)   *I know we all get a bit stuck in our ways once we're **over** thirty, but for Heaven's sake, isn't it worth a try?* (1991, BNC)

---

1 To be sure, *bona fide* PPs can marginally function as subjects and even objects, as in *under the bed is a good place to hide* (Bresnan 1994). However, the use is subject to semantic constraints, working best "in contexts in which the semantics require or the context presupposes a place or time argument" (Bresnan 1994: 110). Moreover, our own exploration of corpus data indicates that the use is decidedly rare and seems to find its natural habitat in pseudo-clefts (e.g. *under the bus is where they belong* (COCA)). For these reasons, we assume no significant connection of this peculiar use to the much more frequent and syntactically distinct adnumeral markers we are concerned with here.

Over thirty in (5) functions as subject complement. As such, it occupies a syntactic slot that would allow both NPs and PPs. Therefore, *over thirty* could be either. Because ambiguity is common today, the shift in *over* from preposition to adnumeral looks very much like a classic example of syntactic reanalysis and actualization. The traditional reanalysis-and-actualization model assumes that ambiguous surface sequences are first assigned a new underlying representation (reanalysis), which subsequently results in manifestly new syntactic behaviour (actualization) (Langacker 1977, Timberlake 1977). On that scenario, examples like (5) could be produced by a speaker who treats *over* as a preposition, but would allow for a hearer to newly interpret *over* as an adnumeral marker (reanalysis). On that same scenario, examples like (3)-(4) illustrate the new syntactic behaviour that becomes possible once *over* has been assigned its new structural status, but that had not been possible under its earlier status (actualization). However, the scenario projected by the reanalysis-and-actualization model, which casts ambiguity as the trigger of syntactic reanalysis, is not borne out by the historical corpus evidence, as we will argue next.

3 DATA

In order to assess the role of ambiguity in the development of adnumeral *over*, we have traced the history of *over* back to its first attested adnumeral uses. As a source of data, we have made use of the (off-line) *Hansard Corpus* (HC), which consists of the transcripts of debates in the British Houses of Parliament, covering the period 1803-2005 and totalling 1.6 billion words. The HC has been chosen for its size and homogeneity, supporting fine-grained and reliable diachronic comparison. Although the HC data has some connection to spoken usage, in the sense that parliamentary records reflect things said in parliament, anyone familiar with the material will recognize that it is a long way removed from spontaneous speech (Mollin 2007). In all events, the corpus represents a highly formal and probably conservative register. However, as shown below, findings appear consistent with data from the smaller but more varied *Corpus of Late Modern English Texts* (CLMET, version 3.1).

From the HC, we have automatically extracted all occurrences of *over* followed by a numerical expression. Judging by frequency alone and prior to any further analysis, the results of the query suggest that adnumeral *over* arose in the first half of the nineteenth century and was fully established by the first half of the twentieth. We have therefore focused the analysis on that period, selecting the data from the 1810s, 1820s, 1830s, 1840s, 1850s, 1870s,

1890s, 1910s and 1930s for closer analysis,[2] with capped sample size at 400 hits per decade (sampled randomly).

In the subsequent manual analysis, we have first filtered out all noise from the data, retaining those instances that could qualify as adnumeral *over*. That requires distinguishing potentially adnumeral from prepositional *over*. To draw the distinction, we have assumed that potentially adnumeral uses will typically allow omission of *over*. In contrast, prepositional uses will not allow omission of *over* because *over* heads a PP, which requires a preposition. For example, omission of *over* is possible in (6) (cf. *those regiments which were 1,000 strong*) but not in (7) (cf. **with a population 4,000*), indicating that *over* is possibly an adnumeral in (6) but certainly a preposition in (7).

(6) *The pecuniary reduction of the Militia force amounted to £50,000, which was effected by reducing those regiments which were **over** 1,000 strong.* (1870, HC)

(7) *The borough of Reigate, with a population a little **over** 4,000, had a constituency, not amounting to 200, and returned one Member.* (1849, HC)

The resultant make-up of the data set is summarized in Table 1. Making up the 'retained' column is the group of potentially adnumeral uses. It is only these instances that have been retained for further analysis.

Decade	Corpus size	Total hits	Sample	Retained
1800s	4,975,287	19	19	2
1810s	7,134,422	20	20	1
1820s	11,613,680	59	59	2
1830s	28,065,927	138	138	3
1840s	30,373,687	259	259	58
1850s	32,992,161	368	368	107
1870s	37,100,705	1,153	400	231
1890s	51,159,886	2,727	400	285
1910s	79,819,189	6,157	400	319
1930s	95,190,137	8,335	400	295

Table 1   Structure of the data set

2 The decision to sample every decade between 1800 and 1850 and only every second decade from 1850 onwards has been made to compensate for the relative sparsity of data in the earlier phase of the development, which is due to the lower frequencies of the target pattern and the smaller size of the corresponding sections of the *Hansard Corpus*.

Within the group of potentially adnumeral instances of *over*, we have further distinguished between ambiguous and unambiguously adnumeral *over* – henceforth labelled simply 'ambiguous' vs 'non-ambiguous'. Instances have been regarded as non-ambiguous if they are the first element of a constituent filling a syntactic slot that resists PPs but welcomes NPs.[3] The majority of instances classified as such are part of NPs functioning as subjects, as direct objects to transitive verbs and as complements to prepositions, as in (3), (4a) and (4b) above. In addition, instances also occur inside NPs functioning as existential subjects, as in (8a), or sometimes as 'genitives of measure' (Quirk, Greenbaum, Leech & Svartvik 1985: 322), as in (8b).

(8) a. *He ascertained that the Dervishes were present in large numbers; there were **over** 2,000 of them.* (1914, HC)

b. *In one Irish county as many as 38 people have received an average of **over** two months' imprisonment with hard labour for the new crime of attending meetings of suppressed branches of the National League.* (1890, HC)

Instances have been regarded as ambiguous if they are the first element of a constituent that occupies a syntactic slot permitting both NPs and PPs. Accordingly, ambiguous *over* occurs as the introducing element of subject complements, as in (5) above, but also in appositions, as in (9a), certain noun postmodifiers, as in (9b), certain adverbials, as in (9c), among other contexts.

(9) a. *no messenger or visitor ever came from the Consulate at Cairo during the whole time (**over** three months) that Mr. Hamilton lay in an almost hopeless condition at Zagazig.* (1874, HC)

b. *I wish to state one or two reasons why the privilege of free education should be extended to children **over** 14 years of age.* (1891, HC)

c. *The Clerk, I am informed, has been in office **over** forty years* (1895, HC)

Where in doubt, we have checked contemporary usage. For example, in nineteenth-century English, noun postmodifiers of the form *over* NUM *years of age*, as in (9b), allow both omission of *over* and replacement of *over* by (decidedly non-adnumeral) *of*, as shown in (10), supporting both the adnumeral and the prepositional reading.

---

3 For simplicity, we assume phrases that consist of only a numeral (and its optional modifiers) are NPs (e.g. *Canetti had read more by the time he was **16** than most of us manage by the time we are **40*** (1989, BNC)).

(10) a. *SIR R. WEBSTER pointed out, to save time, that Clause 3 extended in certain cases to **children seven years of age**.* (1894, HC)

b. *As regards the standards, they are such that **children of 11 years of age may, and do, easily pass the Fifth Standard**.* (1898, HC)

4   Results

So what do these data tell us? Separating out the non-ambiguous adnumeral uses of *over* from the ambiguous uses, Figure 1 shows the frequency of both in the HC throughout the period studied. The full lines show normalized frequencies (per million words), revealing the characteristic slow-fast-slow progression of change, while the dotted line shows the relative share of ambiguous uses over time (as a percentage). Note that given the very low numbers of attestations, figures for the first two decades have been aggregated in a single data point.
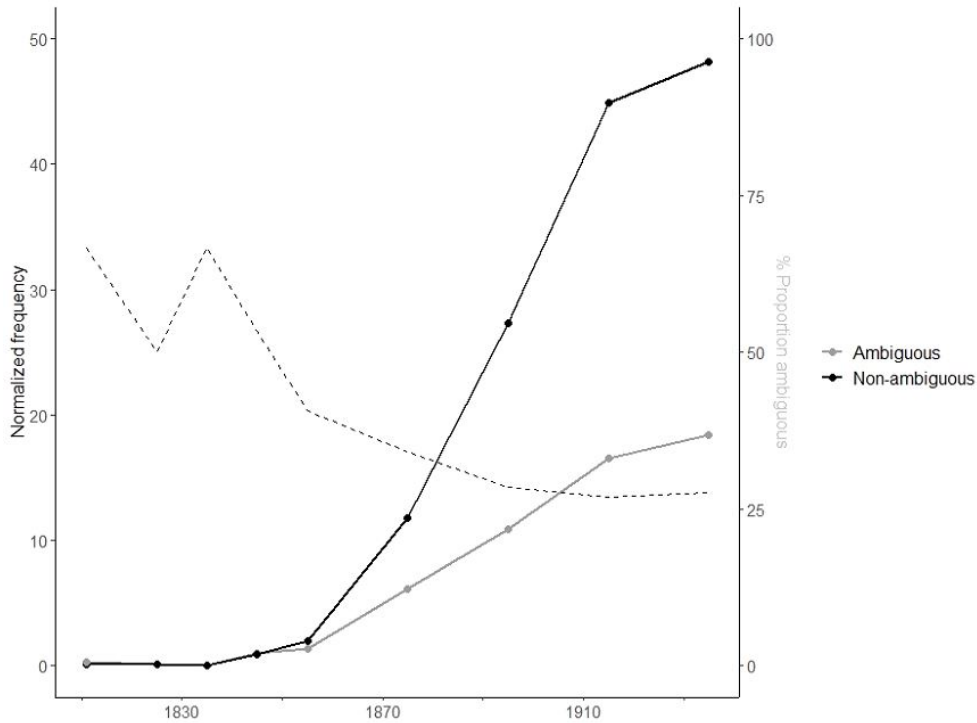

Figure 1   Ambiguous and non-ambiguous *over* in the HC.

One thing that is immediately clear from Figure 1 is that non-ambiguous ad-numeral and ambiguous uses emerged virtually at the same time. Out of the first five potentially adnumeral instances of *over* in the HC, occurring from the 1800s to 1820s, three are ambiguous and two are non-ambiguous, given here in (11) and (12) respectively.[4] The occurrences of *over* in (11) are in ambiguity contexts, as identified above, with *over* introducing a subject complement in (11a) and (11b) and a noun postmodifier in (11c). In contrast, the occurrences in (12) are non-ambiguous adnumerals: *a little over $1^{1/2}d.$* in (12a) functions as subject of an elided clause (with adnumeral *over* premodified here by the degree adverb *a little*), while *over 100 barrels* in (12b) functions as direct object to *brewed*.

(11) a. *Out of 21 towns where the rate was **over** 1s., 20 would receive relief.* (1807, HC)

b. *his Majesty's ministers had paid the utmost attention to proportion the establishment to the effective strength of regiments: When a regiment for instance was under 400, the establishment was fixed at 40[0]; when **over** 400 and under 600 at 600; and so in progression* (1810, HC)

c. *That female slaves **over** twelve years old, should not be publicly whipped under a penalty of 10l.* (1826, HC)

(12) a. *Of this 9d. is for building, and a little **over** $1^{1/2}d.$ for maintenance* (HC, 1807)

b. *Thus, in case he brewed **over** 100 barrels, and not exceeding 1,000 in the year he would have to pay 7l 5s.* (1824, HC)

An additional check on the *Corpus of Late Modern English Texts* (CLMET, version 3.1), which covers also the eighteenth century, has produced two earlier instances than the ones in (11) and (12), one ambiguous and one non-ambiguous, given here in (13) and (14) respectively. That confirms the picture of near-simultaneous emergence.

(13) *Upon this occasion my case was not brought forward, but was suffered to stand **over** six months longer.* (1794, CLMET3.1)

(14) *Produce of the value of **over** a million pounds per annum is now exported from Napier.* (1769, CLMET3.1)

---

4 The abbreviations *l, s.* and *d.* in these examples stand for pounds, shillings and pennies respectively.

While it is more than likely that additional data would produce additional examples of ambiguous as well as non-ambiguous adnumeral *over*, it is also clear that both patterns occurred in the eighteenth and early nineteenth centuries at extremely low frequencies of around 0.1 occurrences per million words or even less. Of course, one or the other must have appeared first in the language but, whichever it was, at the time when non-ambiguous adnumeral *over* made its entrance, ambiguous *over* was extremely infrequent at best.

It is instructive to put these figures in further perspective. Compared to ambiguous *over*, unambiguously prepositional *over* was massively frequent. Estimates of the frequency of prepositional *over* in CLMET and HC indicate that, for every ambiguous instance of *over*, eighteenth and early nineteenth-century language users would have encountered several thousands of unambiguous prepositional uses.[5] That is very far indeed from the roughly 50% bridging uses exposure seen in the changes analysed by Winter-Froemel (this vol.). That is to say, at the time when it mattered, language users had every reason to analyse *over* as a preposition. They would have been unlikely to assign any other structural status to *over* based only on the fleetingly rare occurrence of structural ambiguity. Thus, the emergence of ambiguous *over* did not, in any historically relevant sense, precede or lead to the emergence of adnumeral *over* – if anything, ambiguous and adnumeral *over* emerged as part of the same change. Ambiguity, then, did not play any credible role of significance as a factor triggering syntactic reanalysis.

That said, there is still the development of adnumeral *over* following its first appearance. Looking at the whole development in Figure 1 above, occurrences in ambiguous contexts appear to be proportionately very frequent and especially so in the earlier stages of the development. Ambiguous contexts are seen over time to take up a large – if gradually shrinking – share of the instances of *over*, starting at above half of all occurrences in the early nineteenth century, and still taking up about a quarter by the 1930s. That suggests that, despite its lack of involvement in the very onset of change, ambiguity was somehow relevant to the subsequent development of adnumeral *over*, and more so in its earlier stages than in later stages.

In interpreting this finding, it would be helpful to have a baseline to tell us how many ambiguous and non-ambiguous occurrences one could reasonably expect to find. Because ambiguity arises simply from the selectional properties of the syntactic slots in which *over*-initial phrases are used, it is inevitable that adnumeral *over* will some of the time occur in ambiguous contexts, even if

---

5 CLMET has 4,462 instances of *over* identified by TreeTagger as preposition for the period 1710-1780, amounting to 425.8 occurrences per million words. The pos-tagged version of the HC, available through https://www.english-corpora.org/hansard/, gives a frequency of prepositional *over* around 200 per million words in the first decades of the nineteenth century.

ambiguity were not a causal factor or driving force in its development. Therefore, to better understand the impact of ambiguity on the long-term development of adnumeral *over*, we have tried to assess how much ambiguity one would find anyway – that is, regardless of whether ambiguity had any special role in the historical development of *over*. To this end, we have compared the observed rate of occurrence of ambiguous *over* with an estimate of its expected rate of occurrence.

To do so, we have started from the assumption that *over* can freely occur in any NP with a numeral. From that assumption it follows that, other things being equal, the syntactic distribution of adnumeral *over* can be expected to resemble the syntactic distribution of numerals without adnumeral *over* – henceforth 'bare numerals'. On that logic, the expected rate of occurrence of ambiguous *over* is the rate of occurrence of bare numerals where *over*, were it to be added, would be ambiguous. Consider the following examples. The bare numeral NP *70 per cent of the electorate* in (15a) functions as complement to a preposition. If it were preceded by *over*, *over* would in that context be a non-ambiguous adnumeral (cf. *the vote rose to over 70 per cent of the electorate*). In contrast, the bare numeral *18* in (15b) functions as subject complement. If *over* were to occur in this context, it would be ambiguous between adnumeral and preposition (cf. *whether or not she was over 18*).

(15) a. *Is it not a fact that in the North-West Frontier Province the vote rose to **70 per cent of the electorate**?* (1932, HC)

b. *a young lady was subjected in Sheffield a few weeks ago to cross-examination on the question of whether or not she was **18**, because she happened to be on licensed premises enjoying a glass of port* (1934, HC)

The two types in (15a) and (15b) can be thought of as 'would-be non-ambiguous' and 'would-be ambiguous' respectively. The expected rate of occurrence of ambiguous *over* then corresponds to the rate of occurrence of would-be ambiguous bare numerals to would-be non-ambiguous ones.

To find that rate of occurrence, we have collected bare numerals in the HC, targeting the 1850s (still representing the earlier phase of change in *over* but with enough instances of *over* to draw comparison; cf. Table 1 above) and the 1930s (representing the levelling off of the rise of adnumeral *over*). Following a query for any numeral, instances have been randomly sampled at a 0.2% sampling rate, yielding sample sizes of 850 and 782 respectively. Data has again been manually analysed, whereby would-be ambiguous contexts have been distinguished from would-be non-ambiguous contexts, as explained above. In the process, two types of context have been excluded. Because in our HC data for the period studied adnumeral *over* is never attested

following a determiner (as it is in (4c) above), we have excluded any numeral NPs with an explicit determiner, such as (16a). Similarly, adnumeral *over* never modifies numeral *one* in our HC material, which is why all instances of bare numeral *one*, such as (16b), have likewise been excluded.

(16) a. *the fact remains that we ought not to sacrifice for the trade with **those 22,000 people** the possibilities of the larger trade with **the 1,400,000,000 outside**.* (1932, HC)

b. *if they had acted upon it, the House would never have heard **one word** from the Protestants of Ireland, if Roman Catholics had filled every office in the State.* (1853, HC)

In contrast, we have not excluded contexts where adding *over* to the numeral is merely awkward pragmatically. These include, for instance, contexts involving big unrounded numbers, as in (17a), or cases where the meaning of *over* would run counter the intended message, as in (17b).

(17) a. *he found that the exports had increased from **£78,076,854** in 1852, to **£115,890,857** in 1856.* (1857, HC)

b. *There is a considerable opinion among the medical profession that from the point of view of nursing, **10 days** is not enough.* (1936, HC)

The analysis produces an expected rate of occurrence of ambiguous contexts for *over*, against which the actually observed rate of occurrence can now be compared. This we have done by calculating the natural logarithm of the observed-to-expected ratio of the rate of occurrence for ambiguous contexts. If the observed and expected rates of occurrence are the same, the result is 0 ($= \ln(1)$); if the observed rate exceeds the expected rate, the result is greater than 0; and if the observed rate is below the expected rate, the result will be less than 0. The same can be done for the rate of occurrence of non-ambiguous contexts. The results, both for the 1850s and for the 1930s, have been visualized in Figure 2. They confirm that the use of *over* is skewed towards ambiguous contexts. The effect is very strong in the 1850s, with ambiguous *over* occurring at about three times the expected rate. The effect may even linger on into the 1930s, where ambiguous contexts are still more common than expected. In light of the results in Figure 2, it is very likely that syntactic ambiguity, even if it did not play its expected role, played some role of significance in the development of adnumeral *over*.
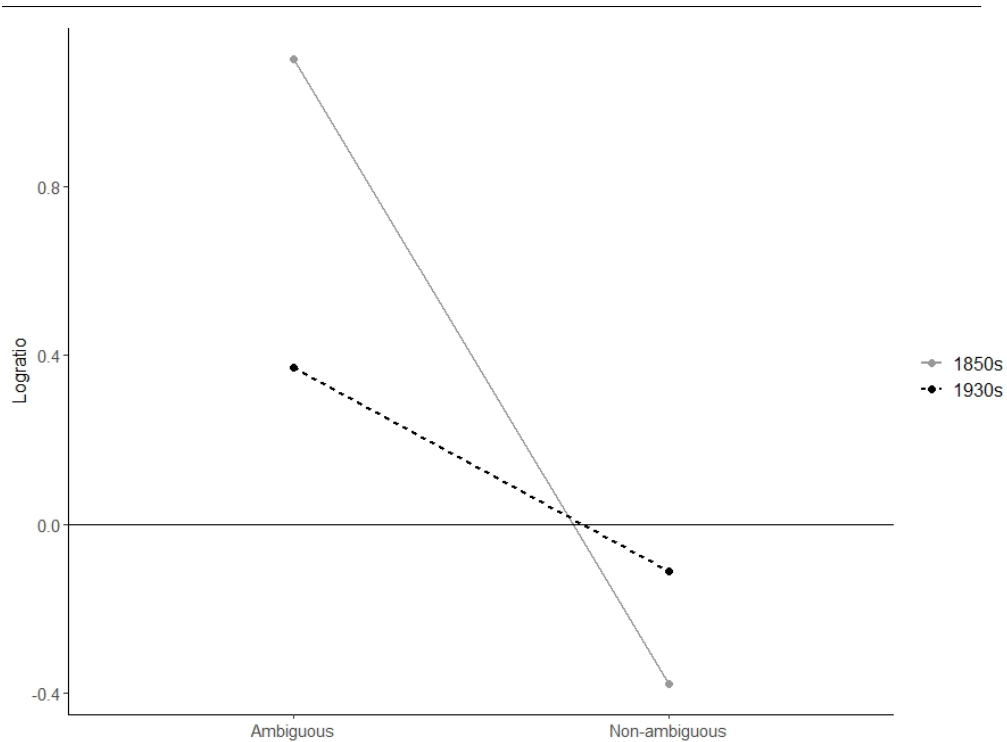
Figure 2   Deviation from expected rates of occurrence for ambiguous and non-ambiguous *over*.

5 AN ALTERNATIVE MODEL OF CHANGE

The above findings are difficult to explain within the traditional reanalysis-and-actualization model. The history of adnumeral *over* shows a very low absolute frequency of ambiguous contexts just prior to the first appearance of non-ambiguous adnumeral *over*. As the first non-ambiguous adnumerals offer an obvious *terminus ante quem* for the moment of reanalysis, we can safely conclude that the occurrence of ambiguous uses was negligible at the time reanalysis could have occurred. At the same time, there is the surprisingly high relative frequency of ambiguous uses once actualization is in progress. All of this runs counter to any predictions the reanalysis-and-actualization model actually makes. First, if the reanalysis from preposition to adnumeral in *over* depended on surface ambiguities, ambiguity should have been prevalent before the appearance of non-ambiguous adnumeral *over*. Second, if actualization were no more than the logical bearing out of the consequences of

reanalysis, then there should be no reason to expect the rate of occurrence of ambiguous contexts to markedly deviate from baseline frequencies during the actualization stage. The findings require another account of syntactic change than the traditional reanalysis-and-actualization model.

To begin with, we have to ask what mechanism could have triggered the shift from preposition to adnumeral in *over*. In fact, the answer is straightforward. *Over* is a recent addition to an established paradigm of adnumeral markers, many of which developed from prepositions and have continued to function as prepositions since. That means the change proceeded along a well-trodden path that is still in evidence synchronically, which makes all the difference. It did not take the confusion caused by ambiguity for speakers of nineteenth-century English to assign a new structural status to *over*. Those speakers had plenty of evidence available to them that prepositions can be repurposed as adnumeral markers. Thus, the change from preposition to adnumeral in *over* is in essence an analogical one (Fischer 2007, De Smet 2009, Aaron 2016), illustrating what De Smet (2009) has called 'categorial incursion', whereby one grammatical category (adnumeral marker) imposes its behaviour on a member of another category (preposition), based on formal and semantic similarity. As Markey & De Smet (2019) show, this at once explains why *over* has been remarkably quick in adopting all the behaviours associated with adnumerals – much quicker than *about*, the oldest member of the paradigm.

The next question is why ambiguous uses nevertheless take up a disproportionately large share of the instances of potentially adnumeral *over*, especially so in the earlier stages of its development. Again, what is involved is probably a form of analogy. De Smet (2012, 2016) has argued that innovations are sneaky, spreading most easily to contexts that resemble established uses. Or, as Israel (1996) put it, "utterances should sound like things the speaker has heard before." In this, superficial similarities may outweigh deeper structural parsing in speakers' assessments of what sounds familiar and therefore feels acceptable (De Smet 2012). It follows that innovations spread faster to contexts where they look superficially familiar, even if they are structurally new.

Ambiguous contexts meet that requirement by definition. Adnumeral *over*, when used in an ambiguity context, can pass for prepositional *over*, which is enough to give it that touch of familiarity that could facilitate its spread. For example, speakers would have been less alarmed by an *over* introducing a subject complement, where *over* has been commonly encountered, than by an *over* following a transitive verb where a direct object is expected and use of *over* is conspicuously unusual. In fact, speakers may even fall short of

activating and selecting *over* for usage in contexts where they have not encountered it before. This way, structural ambiguity provides a perfect cover under which innovations can sneakily spread. At the same time, as the combination of *over* with numerals becomes better entrenched, this creates a new association that increasingly licenses the use of *over* in otherwise unfamiliar contexts.

On this interpretation of events, ambiguity did play a significant role in the history of adnumeral *over*, not as the spark of change but as its fuel. That is, ambiguity did not cause the dependency shift and rebracketing that gave rise to adnumeral *over*. But, when adnumeral *over* was a structural possibility, speakers were more likely to resort to it in ambiguous contexts, because those contexts already had established associations with *over* that were still lacking in non-ambiguous contexts. This interpretation is consistent with the facts.

6   FURTHER CONSEQUENCES

We do not mean to claim that the way adnumeral *over* developed is representative of all cases of apparent syntactic reanalysis and actualization. Nevertheless, lessons can be drawn from the history of *over* that may apply to the way we think of and study reanalysis more generally. We would like to suggest the following take-aways, from less to more speculative.

6.1   *Caution where it comes to ambiguity*

For any change being studied the mere observation of ambiguity in the historical record does not license the conclusion that ambiguity triggered the change. The history of *over* is especially revealing in this respect, because it actually shows a high incidence of ambiguity in the historical record. It takes a large and fine-grained data set to establish that, in absolute numbers, ambiguity was virtually absent in the crucial decades when adnumeral *over* first began to appear. This warrants great caution in interpreting the role of ambiguity in historical change. If we are right to believe that ambiguity can be a factor promoting the spread of an innovation, it is conceivable that successful innovations are especially ones where structural factors support high rates of ambiguity, regardless of whether it was ambiguity that first triggered the innovation or not. In other words, whatever its role, ambiguity is expected to be prevalent in changes, be it as the spark or as the fuel of change, or indeed as both.

To underscore the need for caution, it is worth revisiting here one of the objections raised against ambiguity as a trigger of change in syntactic reanalysis (Fischer 2007, De Smet 2009, 2013). Strictly speaking, structural ambiguity

15

arises when an expression can be assigned to two different possible structural representations. Those structural representations must already exist, as abstract templates associated to different existing classes of expressions. Else, no ambiguity can arise. For example, the first word in [əˈnædər] can be parsed as *a* or as *an*, corresponding to the existing variant realizations of the English indefinite article, leading to divergent parses as either *a nadder* or *an adder* and to the historical change from the former to the latter. It is perfectly plausible that ambiguity triggered change here, but it is only because variation between *a* and *an* is already built into the system of the language that ambiguity could arise in the first place. Now this predicts that reanalysis through ambiguity should be possible in exactly those situations where a categorial incursion scenario is also possible. Again, we do not want to dismiss ambiguity as a potential trigger of change (see below), but we do call for a healthy dose of caution in interpreting its role.

6.2 Types of change

The reader can take the history of *over* simply as a cautionary tale – 'beware of over-applying the reanalysis-and-actualization model!' However, if we really want to learn from the history of *over*, we should also ask how it fits into a broader typology of changes. Without any pretence of exhaustiveness, let us single out two dimensions of interest here.

Along the first dimension, some categorial incursions are likelier than others. At one end, there are category changes that seem to come quite naturally and easily and apparently require no 'bridging uses exposure', to borrow Winter-Froemel's concept (this vol.). Arguably, such changes border on productive processes. The change from preposition to adnumeral in *over* belongs squarely in this category.[6] Why should that be so? As argued above, it undoubtedly matters that *over* entered a neat and well-established paradigm of deprepositional adnumeral markers. Making the paradigm of adnumerals salient as a class is its strict syntagmatic correlates: adnumerals modify numerals. Further, the pathway from preposition to adnumeral is synchronically extant. The other adnumerals that *over* comes to join are all still used as prepositions. Moreover, the share of English prepositions that can be used as adnumerals is relatively high, and so is the share of forms descended from prepositions among adnumeral expressions. Finally, the specific meanings

---

6 The reader may wonder what distinguishes a change from a productive process. We doubt that there is a clear-cut difference, but note that the implementation of adnumeral status for *over*, though decidedly quick, was not so automatic as to be instantaneous. For at least a few decades, adnumeral use lagged behind in specific syntactic environments, such as inside a prepositional phrase or inside a subject phrase.

expressed by adnumerals largely match their prepositional semantics. For example, both as an adnumeral and as a preposition, *over* is an antonym to *under* and a near-synonym to *above*.[7] All of this has probably contributed to a relatively smooth shift from preposition to adnumeral.

At the other end are categorial incursions that appear to be more cumbersome and require an extra nudge – or spark. Ambiguity can be that spark.[8] For example, ambiguous contexts are instrumental in the change from noun to adjective in items like English *key*, as in (18). In (18a), *key* is ambiguous between a noun in a noun-noun compound and a premodifying adjective. In (18b), *key* is an unambiguous adjective. Other items in English that probably underwent the same change include *amateur*, *average*, *bogus*, *niche*, *psycho*, *working class*, etc.

(18)   a. *coal was a **key** factor* (1951, HC)

     b. *good transport is **key** to our economic ambitions* (2000, HC)

For *key*, at least, it has been demonstrated that the ambiguous uses historically precede the emergence of unambiguously adjectival uses (Denison 2001, De Smet 2012) and that their frequency at the level of individual users reliably predicts who is most likely to adopt the innovative adjectival uses (De Smet 2016). Furthermore, a closely related language like Dutch, where for prosodic and morphological reasons the same noun/adjective ambiguity is much more exceptional, sees far fewer changes from noun to adjective (Van Goethem & Smet 2014). As for English, however, the fact remains that the change requires prior ambiguity and even then many seemingly eligible items do not undergo the change or never attain full adjectivehood (e.g. *advance*, *bottom*, *champion*, *surprise*). We suspect that some of the same factors that facilitate the shift from preposition to adnumeral are encumbering the shift from noun to adjective. The historical pathway of change, while synchronically extant in individual items like *key*, lacks salience to language users, because the great majority of adjectives do not have corresponding nouns and vice versa. Adjectives that developed from nouns make up a small and disparate sub-category within the much larger class of adjectives, as do their nominal sources within

---

7 Note, however, that the core member of the adnumeral paradigm, *about*, is losing its connection to its prepositional origin as its prepositional spatial sense 'around' is in the process of becoming obsolescent.

8 For some changes to take place even systematic ambiguity does not suffice and an additional 'spark' is needed. For example, De Smet (2014) shows that the distinction between English gerunds and present participles, both of which end in *-ing*, is remarkably resilient to confusion, even in contexts where the features that overtly distinguish the two categories are rarely manifest. Typically, it appears to take some additional trigger for speakers to reinterpret present participles as gerunds or vice versa.

the class of nouns. Across items that underwent the change, semantic relations between noun use and adjective use are unsystematic, typically involving lexically specific semantic extensions. As a result, the potential for analogical extension is much less obvious to language users. The exceptions prove the rule here. Noun-to-adjective changes are entirely unproblematic (and, arguably, productive) if items belong to a recognizable subclass where double class membership is systematic, such as colour nouns (*apricot*, *burgundy*, *eggshell*, etc.).

Along the second dimension, some changes may be sneakier than others. A well-known view of language change holds that linguistic innovations are motivated by 'extravagance' (cf. Keller 1994 and Haspelmath 1999). Innovative linguistic behaviour, by deviating from norms, draws attention to the speaker and their message and, in doing so, serves a socio-pragmatic purpose on the part of the speaker. At first glance, this is at odds with our own sneakiness claim, which holds that changes are typically inconspicuous. However, the relationship between sneakiness and extravagance is a complex one. Consider the strengthening of historically weak verbs in Germanic languages, where we find both very extravagant and very unextravagant changes. Many instances of strengthening happen unnoticed and are unintentional, driven by analogical pattern-matching. Some instances, however, are tongue-in-cheek and therefore decidedly extravagant. As it turns out, the tongue-in-cheek instances resemble the unintentional ones (Knooihuizen & Strik 2014: 196, I. De Smet 2021: 35). For example, in Dutch, the weak verb *snappen* 'understand' acquired a playful strong past participle *gesnopen* in addition to its weak form *gesnapt*. Even so, the new form *gesnopen* happens to have the vowel change to /o/ that is a dominant pattern in Dutch strong verbs and that is similarly adopted by most other (including unintended) newcomers to the strong paradigm. The reason is that deliberate innovations rely on the same processes of analogical pattern-matching as do unintentional innovations. Extravagance guides speakers to making specific choices from the linguistic options available to them. But those options can only come into being through the cognitive machinery that lets speakers generate linguistic output in the first place, including the sneaky workings of analogy.

What makes extravagant choices stand out, then? While extravagant choices are not without system-internal motivation, they may be internally motivated to a lesser degree. For example, *gesnopen*, despite its /o/ vowel, is no exact fit for any of the Dutch strong verb classes: in particular, the short /ɑ/ in its infinitive stem sits in the way of easy classification (I. De Smet 2021: 87). At the same time, a lack of motivation is not the only, and perhaps not even the main source of extravagance. A choice can be both strongly motivated and

extravagant as long as language users are aware of a more conventional alternative (cf. Goldberg 2006 on 'statistical preemption'): the humour in *gesnopen* is lost on anyone who does not know the conventional weak form *gesnapt*. All this implies that extravagance and sneakiness are partly related, partly independent of one another, and operate at interacting but distinct levels of linguistic processing. Speakers may be attracted to the unexpectedness of an extravagant lexico-grammatical choice while at the same time being oblivious to the often subtle, internally motivated semantic and syntactic extensions that made their choice possible.[9]

We can now specify our second dimension. Just as there are degrees to linguistic creativity (Hoffmann 2018), so there can be degrees to the relative share of sneakiness and extravagance in linguistic innovations. At one end there are the many changes that proceed completely unnoticed. Here, system-internal motivation is maximal, conventional alternatives are lacking and norm-deviation is close to imperceptible. At the other extreme, norm-deviation is blatant, even as some degree of system-internal motivation remains indispensable. In the middle sit innovations that have strong internal motivation but derive extravagance from the existence of a conventional alternative. We expect that that is where the emergence of adnumeral *over* is to be situated. Presumably there was a time when adnumeral *over* was extravagant compared to the conventional alternatives *more than* and *above*, but that does not mean that its use was ever independent of the analogical relations supporting it, be it the paradigmatic relation to other adnumerals, or the surface resemblance to the prepositional uses of *over* in ambiguous contexts.

6.3   *Ambiguity beyond syntax*

Once ambiguity is potentially thought of as the fuel rather than the spark of syntactic change, the question arises whether that way of thinking can be extended to other areas of change, notably semantic change. Tellingly, Traugott argues that even after semantic innovation has created a new grammatical construction, ambiguous contexts can remain relevant as "an essential part of the ecology of [its] generalization and further development" (Traugott 2012: 231). As such, our own position is an echo of Traugott's: ambiguity is potentially more than the first enabler of change and can continue to fuel the rise of a new pattern far beyond the point of its first appearance. More generally, it has already been recognized that semantic changes may need some kind of additional system-internal support to gain a toehold in the language system.

---

9 Our interpretation is, perhaps, not too far removed from Keller's and Haspelmath's original conception. As Keller (1994) makes clear, speakers usually have no actual intention of changing the language. All they want is for their message to stand out.

A highly relevant concept, in this respect, is what Heine (2002: 84-85) has called 'switch contexts'. In Heine's model of semantic change, the shift from a source to a target meaning is made possible by 'bridging contexts' that are semantically ambiguous. However, it is in the special subset of switch contexts that the first effects of semantic change are seen. In such contexts, the target meaning is foregrounded, the source meaning is ruled out or backgrounded, but use of the target meaning is nevertheless still restricted to "the context that gave rise to it". Only later, in the 'conventionalization' stage of change, does the target meaning appear in contexts that are completely novel with respect to the source meaning. According to Heine, switch contexts play a crucial role in mediating the shift from source to target meaning. Without them, the target meaning would remain just an implicature.

It is not too hard to draw an analogy here to the history of *over*. Switch contexts resemble the contexts that adnumeral *over* was initially attracted to most. Those were the contexts that would have been compatible with the old prepositional syntax of *over* as they retained some semblance of the old representation. In other words, a similar role is played by switch contexts in semantic change and the syntactically ambiguous contexts seen at work in the history of *over*. Both types of context heighten the appearance of normality and in doing so facilitate the spread of an innovation, acting not so much as the spark but as the fuel of language change. The least this can be taken to imply is that the caution we called for regarding ambiguity in syntactic change extends to ambiguity in semantic change. Its role may be more complicated than is often assumed.

7 Conclusion

To summarize, in this paper we have drawn a detailed picture of the development of English *over* and have used it to assess the role of ambiguity in the shift from preposition to adnumeral. The traditional reanalysis-and-actualization model of syntactic change predicts that ambiguity should precede the first evidence of syntactic innovation, but that is not what we found. In the history of adnumeral *over*, ambiguity is common, but it is common at the wrong time. We found that ambiguous and non-ambiguous adnumeral instances of *over* emerged simultaneously, with ambiguous uses far too infrequent in absolute numbers to offer a credible basis for ambiguity-triggered reanalysis. At the same time, even with overall low absolute numbers, early usage of adnumeral *over* is characterized by a higher-than-expected relative share of ambiguous instances.

The findings reveal that syntactic change can happen without ambiguity as trigger. The likely mechanism at work in the case of *over* is categorial

incursion, which is an analogy-driven process, whereby one established abstract category takes over from another as the model on which an expression's behaviour is shaped. The shift is motivated by the relative degrees of formal and semantic fit between expressions and categories. The subsequent relative prevalence of ambiguous contexts can be explained by a tendency for innovations to spread more easily to contexts where the innovation is less conspicuous. The tendency is another manifestation of analogy, although at a deeper level it may reflect variation in the ease of activation of an innovative expression, as determined by pre-existing associations between forms and syntagmatic contexts.

The key message from the development of adnumeral *over* is that the relation between change and ambiguity is bidirectional. Ambiguity attracts change, but change also attracts ambiguity. This warrants caution in interpreting the role of ambiguity in any given change, particularly when the historical record is patchy or non-existent. It also opens up new avenues for research, asking which changes are most or least dependent on ambiguity. Finally, it provides ground for speculation about semantic change, where similar effects may well be at work.

Acknowledgements

References

Aaron, Jessi. 2016. The road already traveled: Constructional analogy in lexico-syntactic change. *Studies in Language* 40. 26–62.

Anttila, Raimo. 1989. *Historical and comparative linguistics*. Amsterdam: John Benjamins.

Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 70. 72–131.

Campbell, Lyle. 1998. *Historical linguistics: An introduction*. Cambridge, MA: MIT Press.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.

De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119. 1728–1755.

De Smet, Hendrik. 2012. The course of actualization. *Language* 88. 601–633.

De Smet, Hendrik. 2013. Does innovation need reanalysis? In Evie Coussé & Ferdinand Von Mengden (eds.), *Usage-based approaches to language change*, 23–48. Amsterdam: John Benjamins.

De Smet, Hendrik. 2014. Constrained confusion: The gerund/participle distinction in Late Modern English. In Marianne Hundt (ed.), *Late Modern English syntax*, 224–238. Cambridge: Cambridge University Press.

De Smet, Hendrik. 2016. How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change* 28. 83–102.

De Smet, Isabeau. 2021. *De sterke werkwoorden in het Nederlands. Een diachroon, kwantitatief onderzoek*: Department of Linguistics, KU Leuven dissertation.

Denison, David. 2001. Gradience and linguistic change. In Laurel J. Brinton (ed.), *Historical linguistics 1999*, 110–144. Amsterdam: John Benjamins.

Denison, David. 2017. Ambiguity and vagueness in historical change. In Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*, 292–318. Cambridge: Cambridge University Press.

Diewald, Gabriele. 2006. Context types in grammaticalization as constructions. *Constructions* 1.

Eckardt, Regine. 2011. Grammaticalization and semantic change. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*, 389–400. Oxford: Oxford University Press.

Evans, Nicholas & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76. 546–592.

Fischer, Olga. 2007. *Morphosyntactic change: Functional and formal perspectives*. Oxford: Oxford University Press.

Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Harris, Alice C. 2003. Cross-linguistic perspectives on syntactic change. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 529–551. Oxford: Blackwell.

Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37. 1043–1068.

Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 83–101. Amsterdam: Benjamins.

Hock, Hans Henrich. 1986. *Principles of historical linguistics*. Berlin: de Gruyter.

Hoffmann, Thomas. 2018. Creativity and construction grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66. 259–276.

Israel, Michael. 1996. The way constructions grow. In Adele Goldberg (ed.), *Structure, discourse and language*, 217–230. Stanford: CSLI.

Keller, Rudi. 1994. *Language change: The invisible hand in language*. London: Routledge.

Knooihuizen, Remco & Oscar Strik. 2014. Relative productivity potentials of Dutch verbal inflection patterns. *Folia Linguistica Historica* 35. 173–200.

Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 57–139. Austin: University of Texas Press.

Markey, Marie-Anne & Hendrik De Smet. 2019. The speed of change: Is grammaticalization faster when it has an analogical model? Paper presented at SHES, Edinburgh, 18–19 May 2019.

Mollin, Sandra. 2007. The Hansard hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora* 2. 187–210.

Plank, Frans. 2004. Inevitable reanalysis: From local adpositions to approximative adnumerals, in German and wherever. *Studies in Language* 28. 165–201.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Timberlake, Alan. 1977. Reanalysis and actualization in syntactic change. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 141–177. Austin: University of Texas Press.

Traugott, Elizabeth C. 2012. On the persistence of ambiguous linguistic contexts over time: Implications for corpus research on micro-changes. *Language and Computers* 75. 231–246.

Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Van Goethem, Kristel & Hendrik De Smet. 2014. How nouns turn into adjectives: The emergence of new adjectives in French, English and Dutch through debonding processes. *Languages in Contrast* 14. 251–277.

Winter-Froemel, Esme. 2014. Re(de)fining grammaticalization from a usage-based perspective: Discursive ambiguity in innovation scenarios. *Folia Linguistica* 48. 503–556.

Hendrik De Smet
University of Leuven
Department of Linguistics
Blijde-Inkomststraat 21, box 3308
B-3000 Leuven, Belgium
hendrik.desmet@kuleuven.be
https://www.arts.kuleuven.be/english/
research/researchers/hendrikdesmet

Marie-Anne Markey
University of Leuven
Department of Linguistics
Blijde-Inkomststraat 21, box 3308
B-3000 Leuven, Belgium
marieanne.markey@kuleuven.be
https://www.arts.kuleuven.be/ling/
func/members/marie-anne-markey