

From *twig-skinny* to *Kate Moss skinny*: expressing degree with common and proper nouns¹

TURO VARTIAINEN

University of Helsinki

¹ I would like to offer my sincere thanks to the two anonymous reviewers for their constructive feedback. I also thank Jukka Suomela and Tanja Säily for methodological discussions. All remaining errors are, of course, my own. I also acknowledge the generous funding by the Academy of Finland grant 276349 to the project “Reassessing language change: the challenge of real time”.

Abstract

This paper provides a constructional (CxG) analysis of N-ADJ compounds in which the noun receives a degree reading (e.g. *bullet-straight*, *Kennedy-handsome*). A semantic analysis based on similes and scale matching is provided, and the recent history and increased productivity of the construction are examined in light of data from both the *Corpus of Historical American English* and a range of present-day corpora. The article introduces new evidence of the increased functional flexibility of both common and proper nouns in English and discusses the ongoing conventionalisation of proper noun degree modifiers both in American English and in other varieties of English. The results of the study suggest that the recent introduction of proper noun degree modifiers has been supported both by constructional (semantic) change and by macro-trends that have affected English usage more generally.

Keywords: Degree modifiers, nouns, compounding, productivity, Construction Grammar

1 INTRODUCTION

In his monograph on English word formation, Marchand (1960: 47) pays attention to an interesting compounding pattern where a noun invites a degree reading. He provides a long list of examples, including compounds like *grass-green*, *ice-cold*, *honeysweet*, *milk-white* and *coal-black*. In all these cases, the noun could be replaced by a more regular degree modifier, such as *very* or *extremely*, without a significant change in meaning (see also Plag 2003: 152; Günther et al. forthcoming). As Marchand points out, the scalar reading in these compounds typically implies a very high degree (see also Bauer 2017: 100). For example, if something is described as *grass-green*, its colour is interpreted to be a sort of deep and intense green, whereas someone who is *whip-thin* is regarded as extremely thin. Typically, the association between the property denoted by the adjective and the one associated with the noun is motivated, as in all the examples above. However, there are also compounds in which it is much more difficult to see how any of the attributes that are associated with the noun could match with the property denoted by the adjective (e.g. *pig-drunk*, *piss-poor*, *dirt-poor*, *stone-blind*, *dog-cheap*).

The N(Deg)-ADJ² compounding pattern is an old one, and some of the forms, such as *brimcald* ('ocean-cold'), *brandhata* ('fire-hot') and *goldbeorht* ('gold-bright'), are already attested in Old English (Chapman & Christensen 2007: 451–452). However, as Marchand (1960: 47) observes, most of the compounds that are used in PDE first appear in the Modern period, and the pattern has remained productive up to the present day (e.g. Lipka 1966, Marchand 1960; also Norrick 2010). However, in spite of this

² This shorthand is simply intended to capture the idea that the construction consists of a noun that functions as a degree modifier and an adjective. I will later use CN(Deg)-ADJ and PN(Deg)-ADJ to denote patterns with common nouns and proper nouns, respectively.

increased productivity, the pattern has remained relatively resistant to systematic corpus-based investigation. In his article on comparative compounds, Norrick (2010) notes that compounds of the type *sky-blue* and *lightning-fast* are not frequent enough to warrant corpus-linguistic exploration, and he therefore focuses his examination on the most frequent comparative compounds in his corpus (i.e. compounds in *-shaped*, *-sized* and *-colored*). Indeed, the small size of corpora has often been quoted as an obstacle for corpus-based research of productivity, and this problem affects diachronic research in particular (Cowie & Dalton-Puffer 2002; Säily 2014). However, the amount of data that we have at our disposal has significantly increased in recent years, thus greatly facilitating the study of low-frequency items and their productivity from both synchronic and diachronic perspectives. For instance, the recently published *iWeb Corpus* includes 14 billion words of text, while the *Corpus of Historical American English* (COHA), which comprises over 400 million words from 1810 to 2009, has been extensively used to study language variation and change since its publication (Davies 2010).

It is therefore now possible to study the history of the N(Deg)-ADJ compounding pattern from a corpus linguistic perspective. We will see that the common noun pattern, which has its roots in Old English, has recently been supplemented by a pattern where proper nouns are used to express degree (e.g. *Mick Jagger thin*, *Han Solo cool*). In order to fully understand the emergence and development of the proper noun construction, we need to take into account both the diachrony of the common noun pattern and the more general trends affecting (proper) noun usage in English. On the one hand, then, we see a micro-change affecting a single construction (N(Deg)-ADJ). On the other hand, the use of nouns as degree modifiers can be understood more generally as part of the increasing

functional flexibility of both common and proper nouns in English (e.g. Rosenbach 2007, Breban 2018, the contributions in this special issue).

This paper is organised as follows. In section 2, I will provide a brief semantic and constructional analysis of the N(Deg)-ADJ compounds studied. Section 3 introduces the corpora used in the empirical case studies as well as the principles of data inclusion on the basis of the discussion in section 2. Section 4 discusses the results of a diachronic study of N(Deg)-ADJ compounds in the *Corpus of Historical American English* and provides some comments on the global diffusion of the PN(Deg)-ADJ pattern, as well as examples of recent innovative usage. Section 5 concludes the paper with a summarising discussion and suggestions for future research.

2 NOUN-ADJECTIVE COMPOUNDS EXPRESSING DEGREE

2.1 *N(Deg)-ADJ compounds as non-literal comparisons*

The kinds of noun-adjective compounds that are studied in this paper are exemplified in (1) and (2) (from the *Corpus of Contemporary American English* and the Internet).

- (1) A pop guitarist named Dick Dale had earlier pioneered instrumental surf music, capturing the energy of the waves in **lightning-fast guitar arpeggios** and wails, but the Beach Boys were the first band to spin beach life into song. (COCA, MAG, 2004)
- (2) Why it is that the most aggressive dangerous dogs, in my experience, always tend to be the most mentally brilliant dogs, I don't know what that's about, but Bandit was **this incredibly brilliant, Einstein smart dog** who would just attack people.

(<http://dogsplayingforlife.com/animal-wise-radio-transcript-aimee-sadler/>; accessed May 24, 2018)

I would suggest that (1) and (2) can naturally be analysed as condensed similes, that is, non-literal (metaphorical) comparisons between two entities. In (1), the target of the metaphor, *guitar arpeggios* is in part understood in terms of the properties of the source, *lightning*, while in (2), the intellect of *Bandit the dog* is understood in terms of *Einstein*. According to this view, these compounds are roughly synonymous with *the guitar arpeggios were as fast as lightning* and *Bandit was as smart as Einstein*, respectively. Figure 1 describes the correspondence between regular similes (on the right) and condensed similes expressed in the N(Deg)-ADJ construction.³

<i>lightning</i>	<i>fast</i>	<i>guitar arpeggios</i>	↔	<i>Dick Dale</i>	<i>is as fast as</i>	<i>lightning</i>
<i>Einstein</i>	<i>smart</i>	<i>dog</i>		<i>Bandit</i>	<i>is as smart as</i>	<i>Einstein</i>
SOURCE	TERTIUM	TARGET		TARGET	TERTIUM	SOURCE

Figure 1. *Lightning-fast guitar arpeggios* and *an Einstein-smart dog* as condensed similes.

Similes have received a fair amount of attention in psychological and linguistic research, and they can be analysed from many perspectives. First, similes can be understood in terms of comparison/contrast (Tversky 1977). According to this view, language users compare the properties of the source and the target and are able to decode the meaning of the expression when matching properties are found. As Ortony

³ Israel et al. (2004: 129) argue that a simile must be overtly marked by a word such as *like* or *as*. However, they do not take the possibility of a condensed construction like the N(Deg)-ADJ into consideration.

(1979) points out, these properties may not be of equal importance or equally salient. For example, our view of Albert Einstein largely rests on his superior intellect, which helps language users decode the meaning of an expression like *Einstein-smart*. An alternative way to analyse similes is through class inclusion. In this model, the comparison between the source and the target takes place on a superordinate level. According to the class inclusion theory, the simile is decoded by forming superordinate categories from which language users select the one that provides the best match to the expression in a given context (Glucksberg 2001: 41). For example, *a shark lawyer* can receive at least two interpretations: i) a lawyer who represents a wildlife foundation, and ii) a greedy and predatory lawyer (see Goldvarg & Glucksberg 1998). In the first case, *shark* is understood to refer to the basic level entity, and no metaphorical reading arises as a consequence. In the second case, by contrast, the metaphorical reading of *shark* as a vicious, predatory and opportunistic animal becomes available when language users search for a suitable meaning on a superordinate level in an effort to understand how sharks and lawyers could possibly be alike.

2.2 Scale matching

As presented, the nouns in the N(Deg)-ADJ pattern typically express high degree. In other words, the nouns in N(Deg)-ADJ constructs function as *boosters* (see e.g. Paradis 2000), and in most cases they can be replaced by a more conventional degree modifier with little change in meaning. More specifically, *extremely* (or *very*) could be used in place of the noun in N(Deg)-ADJ constructs like *silver-bright*, *ivory-pale*, *moon-barren* and *corpse-cold* to indicate, somewhat vaguely, that the adjective should occupy a very

high point on a degree scale. Occasionally, however, the typical meaning of the construction can be manipulated, and the noun can express more precise points on the scale. For example, in (3), the playing skills of three ice hockey players are compared to each other. We end up with an ordered list of the players' playing ability, proceeding from two very good players (*Tanev* and *Hjalmarsson*) to an even better one (*Duncan Keith*). Nevertheless, in both *Hjalmarsson good* and *Duncan Keith good* we are dealing with boosters: the first phrase corresponds to *very good*, while the second one corresponds to *extremely good*.

- (3) Tanev is very good **Hjarlmasson** [sic] **good** though not **Duncan Keith good** (curse his name). (iWeb Corpus)

One way to analyse how language users decode this kind of scalar meaning is through scale matching across two domains. In order to understand the degree reading evoked by the source (e.g. *lightning*), language users need to be able to locate its referent on a degree scale and then provide a match for it on a scale on which the target (e.g. *guitar arpeggios*) is situated (see Norrick 2010: 219–220). In the case of *lightning-fast*, the first scale locates *lightning* at the top end on a scale that measures the speed at which natural phenomena can occur, while the second scale measures the speed at which *arpeggios* can be played on the guitar (from very slow to extremely fast). The correct reading arises through the matching of these two scales, and although the values for speed, and even the way speed is measured, differ greatly for music and a natural phenomenon like lightning, the process of scale matching helps us understand that the *arpeggios* are intended to have a very high value for speed.

Here, a remark concerning the difference between literal and non-literal comparisons is in order. For literal comparisons, there is obviously no need to posit two scales. For instance, in *a waist-high fence* the height of both *waist* and *fence* are measured on the same scale, and a statement like *this is a waist-high fence* has a measurable truth value. Sometimes, however, the difference between literal and non-literal comparisons can be extremely subtle. Consider (4a) and (4b), for instance.

(4) (a) Niels Bohr was **as smart as Albert Einstein**.

(b) Ethan Canin. It's not enough that **Canin is Einstein-smart** (he attended Stanford University, the University of Iowa Writer's Workshop and then Harvard Med School), but he's also a lyrical, almost hypnotic, writer.

(<https://www.pdxmonthly.com/articles/2009/10/7/wordstock-culture-100709>)

In (4a) we make a factual statement, and a literal comparison, about how smart Niels Bohr and Albert Einstein were. In a case like this, it suffices to imagine one scale on which the two great minds would be ranked, and the clause will have a measurable truth value. Although the given-new structuring of information in the clause certainly affects the meaning in the sense that here it is Einstein who is used as a measuring rod for intellectual genius, from a purely truth-conditional perspective the statement is reversible (see Glucksberg & Keysar 1990). If the proposition is true, we could also say that *Albert Einstein was as smart as Niels Bohr*.

By contrast, in (4b), the comparison is non-literal, and the statement only concerns someone called Ethan Canin – Einstein's name is only evoked because he is perceived to embody the essence of smartness (see e.g. Searle 1993: 92–93; Glucksberg 2001: 41). Again, it will be useful to assume the existence of two scales. The first scale

can be regarded as consisting of all the people in human history, who are then ranked on a scale according to their intellectual capacity. The second scale consists of people who one is bound to meet at some point in one's lifetime. It is on the first scale that Einstein ranks the highest, while the second scale hosts Ethan Canin at the top end (for the person who made the claim).⁴ So, while both scales measure human intellect, they are not the same scale.

A crucial point for both CN(Deg)-ADJ and PN(Deg)-ADJ compounds concerns the correct identification of the source (cf. Norrick 2010: 216). When it comes to common nouns, the source must be something that is not only immediately identifiable but is also a suitable exemplar for the matched property. This allows for much variation in linguistic usage, as our everyday lives provide us with a rich supply of potential metaphoric sources for various purposes. For example, in the *Corpus of Historical American English* something 'very/extremely bright' is described as *dew-bright, sky-bright, water-bright, mirror-bright, jewel-bright, crystal-bright, silver-bright* and *gem-bright*. By contrast, when a proper noun is used to modify an adjective, we do not have as much choice. Typically, the source in a PN(Deg)-ADJ compound needs to be a "paragon" (Lakoff 1987: 87–88) or an "allusive proper noun" (Antonopoulou 2004: 220), that is, someone who is particularly famous for possessing (or being associated with) a certain property. The set of possible paragons varies, of course, from time to time and culture to culture, which is typical of *pragmatic scales* more generally (see e.g. Levinson 2000, Coulson 2001, Claridge 2011). For instance, in an Anglo-American context it is easy to find examples of incredibly rich people described as *Bill Gates rich*,

⁴ In studies of hyperbole (Fogelin 1988: 13; Claridge 2011: 12), it has been noted that hyperbolic expressions tend to be "corrected away from the extreme". This observation fits well with the idea of scale matching in non-literal comparisons, as the scale for the metaphorical source expresses a higher value than the scale for the target.

Warren Buffet rich or *Mark Zuckerberg rich*, but you do not generally find descriptions like *Amancio Ortega rich* or *Carlos Slim rich* even though both men are included in the top ten of the Forbes 2018 billionaire list. Similarly, Fauconnier’s (1975: 353) example of a “pragmatic superlative” in *Onassis couldn’t afford this place* (to mean ‘nobody could afford this place’) already feels outdated, as Aristotle Onassis is no longer a current paragon in this frame.

2.3 *N(Deg)-ADJ compounds as constructions*

The N(Deg)-ADJ compounds studied in this paper show properties that make them well-suited for a constructional (CxG) analysis. In Construction Grammar, constructions are considered to be more than just the sum of their parts – they carry meaning of their own and may impose constraints on the usage of the constructs sanctioned by them (e.g. Lakoff 1987; Goldberg 1995; Croft 2001). When considering the N(Deg)-ADJ compounds, it seems that constructional semantics would indeed form a desirable part of analysis; after all, nouns that are used in these constructions do not in general denote scalar meanings. To illustrate the benefits of a constructional analysis, let us once again consider (4a) (repeated here as (5a) for convenience), and compare it with (5b).

- (5) (a) Niels Bohr was **as smart as Albert Einstein**.
 (b) Niels Bohr was **Einstein-smart**.

In my view, the difference in the meaning of (5a) and (5b) boils down to constructional semantics: in (5b), the proper noun *Einstein* is coerced into a degree modifier construction, and it consequently inherits the degree meaning from the parent

construction.⁵ In contrast to more conventional degree modifiers like *very* or *extremely*, however, language users may continue to associate the noun in an N(Deg)-ADJ construct with its original referent. For instance, in many N(Deg)-ADJ compounds that express colours of high chromatic intensity (e.g. *grass-green*, *ruby-red*, *sky-blue*, *emerald-green*, *snow-white*), the degree reading becomes intertwined with the qualitative aspects of the nominal referents, and the noun can also be interpreted as a submodifier to the adjective. It can be therefore argued that *emerald-green* is not completely synonymous with *grass-green*, even though both nouns can be used to emphasise the intensity of the colour, as in (6) and (7).

(6) His **grass-green eyes** looked candidly into MacMaine's own. (COHA, Fiction, 1961)

(7) Her skin was pale and **her emerald-green eyes** weren't flashing their usual fire. (COHA, Fiction, 2003)

Examples (6) and (7) are also illustrative because they demonstrate the importance of studying N(Deg)-ADJ compounds in context. While I would argue that both *grass-green* and *emerald-green* in (6) and (7) express degree (or are ambiguous between degree and submodifier readings), it is easy to come up with contexts in which they might only be interpreted as submodifiers (e.g. *I need grass-green paint, not emerald-green*).

This kind of ambiguity typically arises with CN-ADJ compounds denoting colours and physical appearance, but proper noun compounds may also be ambiguous.

⁵ In many strands of Construction Grammar, meaning relations between constructions are described in terms of hierarchical networks, where more concrete constructions inherit meanings from more abstract constructions (e.g. Croft 2001: 25; Goldberg 2003: 222–223; Trousdale 2013). *Coercion* refers to a situation where a linguistic item is used in a construction in which it is not typically used, and its meaning is *coerced* to correspond with the meaning of the parent construction (see e.g. Michaelis 2002; Lauwers & Willems 2011).

For instance, in addition to expressing degree, a compound like *Salieri-mad* can be used to express a particular kind of obsessive and jealousy-induced madness that is associated with Mozart's rival Antonio Salieri in popular writing (see Bybee 2010: 91). Likewise, if a woman is described as *Audrey Hepburn pretty*, it might seem strange if the person in question was a curvaceous blonde instead of a slender brunette.

I would suggest that these ambiguities are a consequence of constructional coercion: when decoding condensed compounds like *emerald-green eyes*, language users may try to find a match not only between the properties of the referent of the coerced noun and the adjective (which would support the degree reading) but also with the referent that is described by the N-ADJ construct (which would support the submodifier reading). If the two referents resemble each other physically, the submodifier reading will be available in addition to the degree reading. By contrast, if the referents do not resemble each other sufficiently (e.g. *mountain-quiet words* or *laser-sharp mind*), only the degree reading will be available. In sum, according to this analysis, constructional coercion is the mechanism underlying the formation of N(Deg)-ADJ constructs, and it is also responsible for the potential ambiguities between degree and submodifier readings. This ambiguity could formally be captured by polysemy links between the N-ADJ structure and the submodifier and degree meanings associated with that structure (Goldberg 1995: 75–77; Hilpert & Diessel 2017: 59).⁶

However, if a coercion-based analysis of N(Deg)-ADJ compounds is accepted, it raises some further questions concerning their status. By definition, coercion implies

⁶ As pointed out by an anonymous reviewer, there are other kinds of N-ADJ compounds, such as *book-smart* and *street-smart*, which are semantically distinct from these two types. The point here is that the ambiguity between degree and submodifier readings arises systematically with certain kinds of adjectives. A more comprehensive analysis all constructions that are formally of the N-ADJ type is outside the scope of this paper.

that a linguistic item is used in a way that differs from its typical usage, and this might indicate that the N-ADJ compounds examined in this paper should not be assigned constructional status. However, there is evidence to suggest otherwise. First, in addition to the degree meaning associated with the compounds, both the CN(Deg)-ADJ and the PN(Deg)-ADJ patterns are subject to constraints that are not inherited from their parent construction: the noun slot must be filled with an exemplar or a paragon for the construction to be used felicitously. There are, admittedly, some nouns which do not follow this constraint. The meaning of these nouns has become bleached in this construction, and they are typically associated with adjectives denoting negative properties (e.g. *dirt-poor*, *piss-poor*, *dog-cheap*). However, these nouns are very few in number, and they should be regarded as exceptions to the general constraints affecting the N(Deg)-ADJ compounds.

A second piece of evidence concerns the CN(Deg)-ADJ pattern in particular. As will be shown in section 4, the common noun pattern has become increasingly frequent and productive in Present-day English. Although there is no consensus on what exactly constitutes sufficient frequency for a form-meaning pairing to be considered a construction (Bybee 2013; Traugott & Trousdale 2013; Hilpert & Diessel 2017), a high frequency of use is generally regarded as evidence for constructional status (Croft 2001: 72). The increased frequency of the CN(Deg)-ADJ pattern would therefore support the claim that we are dealing with a construction in the CxG sense. However, when it comes to the PN(Deg)-ADJ pattern, frequency data point to another direction. Here, the very low frequency of the pattern suggests non-constructional status, and this view gains some support from speaker intuitions. While some native speakers whom I have consulted consider the proper noun pattern (e.g. *Mick Jagger thin*) to be perfectly

acceptable, others have expressed some reservations, pointing out that they would only expect such constructs to be used if the speaker tried to be particularly clever or expressive.

Nevertheless, there is some evidence that, at least for some speakers, the PN(Deg)-ADJ pattern has already become entrenched. This evidence comes from humorous uses of the pattern, where the constructional meaning is manipulated for comic effect. The logic here is simple: in order for people to notice this manipulation, they need to be aware of the constructional meaning, which implies at least a certain degree of conventionalisation.⁷ Example (8) is taken from Eileen Rockefeller's 2013 autobiography, *Being a Rockefeller, Becoming Myself: A Memoir*. Here, Rockefeller plays with the paragon status of her family name in the expression *Rockefeller-rich* and the actual state of affairs in the 1960s, creating a mismatch between the non-literal comparison expressed by the PN(Deg)-ADJ construction and the possibility to interpret the phrase as a literal comparison.

- (8) In the sixties it was embarrassing to be rich. It was even more embarrassing to be
“Rockefeller rich.”

Examples (9) and (10), on the other hand, come from an episode of the animated TV-sitcom *The Simpsons*, which originally aired in 1995. In (9), a Hollywood director promises to make one of the characters in the show, a boy called Milhouse van Houten, as famous as Gabby Hayes, or *Gabby Hayes big*, while in (10), the same director compliments Milhouse by saying that he is *Van Johnson good*. In order for the jokes to be funny, the audience needs to be familiar with the meaning of the construction, which

⁷ To be more precise, this is evidence of cognitive entrenchment, but the fact that these texts are aimed at a large audience suggests that the authors and the copyeditors/producers had considered the pattern to be sufficiently conventionalised.

places the intensified property at the (near-)extreme of the degree scale: although Gabby Hayes had a successful career playing supporting roles in B Westerns, and Van Johnson could even have been called a film star at some point in his career, both actors seem to fall short of the expectation that the proper noun in the PN(Deg)-ADJ construction must be a paragon (some contemporaries of Gabby Hayes and Van Johnson include, for example, Clark Gable, Humphrey Bogart, James Stewart and John Wayne, who would arguably have been better sources for the simile if no humour was intended).⁸

(9) That Milhouse is going to be big. **Gabby Hayes big!**

(10) Milhouse. Listen, you can't quit this movie. I've seen your work. It's good. Very, very good. **Van Johnson good.**

Examples (9) and (10) also show that language users need to have a good understanding of pragmatic presuppositions (see Lambrecht 1994: 52) that constrain the use of PN(Deg)-ADJ compounds in particular. These presuppositions can be used to establish interpersonal rapport between discourse participants. For instance, the writer of a film blog who describes a character in a film as *Sarah Palin stupid*⁹ must have made the determination that a sufficiently large proportion of his readers agree with the proposition that “Sarah Palin is extremely stupid” (or at least understand why such presupposition can be made) before publishing the review. The existence of pragmatic presuppositions like these may also in part explain why some of my informants commented on the expressive quality of PN(Deg)-ADJ compounds.

⁸ Humour arising from the manipulation of scalar meanings has also been discussed in Bergen & Binsted (2004), who analyse humorous and non-humorous uses of the X IS SO Y THAT Z construction (e.g. *it was so cold in the kitchen that there was frost on the lettuce* vs. *yo' momma's so old she was a waitress at the Last Supper*).

⁹ <http://www.thefilmchair.com/wordpress/index.php/2009/10/12/movie-review-paranormal-activity/> Accessed January 15, 2019.

In sum, there are clear constraints pertaining to the use of the two N(Deg)-ADJ patterns which do not concern the degree modifier construction more generally. As discussed, these constraints have to do with encyclopaedic knowledge and speaker/hearer alignment and are therefore pragmatic in nature. Although there is still some debate concerning the status of pragmatic information in constructions (see e.g. Kay 2013 and Cappelle 2017 for differing views), I would argue that the meaning and use of the N(Deg)-ADJ pattern cannot be fully described without recourse to pragmatic information, and I see no reason why this kind of information should not be included in the constructional specification. In short, I would suggest that at least the CN(Deg)-ADJ pattern could be considered a (meso-)construction in its own right: it inherits its general meaning from the degree modifier construction, but its use is subject to specific pragmatic constraints. The degree of conventionalisation of the PN(Deg)-ADJ pattern is more debatable, and while the pragmatic constraints associated with PN(Deg)-ADJ compounds seem to be even stricter than those concerning the common noun pattern, its low frequency, which is discussed in more detail in section 4.3, suggests that it has not yet become conventionalised on a larger scale. I will return to this point in sections 4 and 5.

3 CORPORA AND DATA COLLECTION

The main goal of this paper is to shed more light on the recent history of the N(Deg)-ADJ construction. The bulk of the data are therefore taken from the largest historical corpus of English available today, the *Corpus of Historical American English* (COHA). COHA provides coverage for a two-hundred-year period from 1810 to 2009. The genre

balance of COHA remains relatively stable over time, with fiction encompassing roughly 50 per cent of the data for all periods. Precise figures are available at the corpus website (<https://corpus.byu.edu/coha/>).

As the PN(Deg)-ADJ construction is so rare, the data taken from COHA are supplemented by examples from other corpora as well as the Internet. The corpora consulted include the *Corpus of Contemporary American English* (COCA), the *Corpus of Global Web-based English* (GloWbE) and the *iWeb* corpus. The data from GloWbE allow us to reach some preliminary conclusions about the conventionalisation and global diffusion of the PN(Deg)-ADJ pattern, while the *iWeb* corpus is used to provide examples of recent, innovative usage of the kind that is not well represented in the smaller corpora. Table 1 provides a list of the corpora used in the case studies.

Table 1. Corpora used in the case studies.

Corpus	Variety	Temporal coverage	Size
Corpus of Historical American English	AmE	1810–2009	c. 406 million words
Corpus of Contemporary American English	AmE	1990–2017	c. 570 million words
Corpus of Global Web-based English (GloWbE)	20 varieties of English from around the world	Present day	c. 1.9 billion words
<i>iWeb</i>	Mainly US, GB, CA, AU and NZ	Present day	c. 14 billion words

Although these corpora provide ample material for the investigation of the N(Deg)-ADJ construction, there are some severe challenges concerning data collection. First, using POS annotation in the queries is very problematic in terms of precision: the overwhelming majority of N-ADJ combinations in the corpus data are not of the

relevant type. Consider examples (11a) to (11i), which show a variety of N-ADJ compounds/sequences that need to be weeded out from the results (this list is by no means exhaustive).

- | | |
|---------------------------------|--|
| (11) (a) an iron-rich diet | ‘rich in iron’ |
| (b) a fever-thin appearance | ‘thin from fever’ |
| (c) snow-heavy boots | ‘heavy with snow’ |
| (d) waist-high walls | literal comparison |
| (e) the soap-fat man | ‘a man who sells fat for making soap’ |
| (f) silk-fresh hair | ‘silky and fresh’ |
| (g) a Hollywood tough guy | ‘an actor who is regularly cast as a tough guy in Hollywood films’ |
| (h) that Harvard cool | ‘coolness associated with those who studied at Harvard’ |
| (i) Gore Vidal witty one-liners | ‘witty one-liners by Gore Vidal’ |

So, a basic POS-based search fails because of extremely low precision. It is necessary, then, to come up with a query, or a set of queries, that would have higher precision even at the cost of lower recall. In this study, I decided to search the corpus in an iterative fashion, starting out with twenty common nouns that I knew were used in the N(Deg)-ADJ construction, and which could moreover combine with different kinds of adjectives.¹⁰ After performing the first query on COHA, I did another query with the

¹⁰ These nouns were *brick, bull, cat, dirt, dog, eagle, feather, fox, ghost, honey, iron, lightning, mountain, rock, snake, spider, star, steel, stone* and *sugar*.

adjectives that were used with these twenty nouns in the N(Deg)-ADJ construction with the assumption that if these adjectives are modified by one noun in the construction, they might also be modified by others. The following queries were carried out according to the same principle, alternating between nouns and adjectives in each query, until new types could no longer be found. Although relatively labour-intensive, the benefit of this approach was that I was able to read every individual N-ADJ compound in context and could therefore be satisfied that all the compounds included in the final analysis were of the relevant type.

I first searched the corpus by looking for hyphenated constructs with the query *-ADJ (where ADJ stands for an individual adjective) and N-* (where N stands for an individual noun). After running through all the iterations of the query, which were five in total, I looked for unhyphenated forms based on the list that I had collected. While this last step did not, of course, yield new types, it ensured that the token frequencies were not affected by spelling conventions. In all, this data collection method yielded a total of 9,268 N(Deg)-ADJ tokens, 1,230 N(Deg)-ADJ compounds, 586 noun types and 156 adjective types.¹¹

The method used in data collection was designed to strike a balance between reasonable recall and amount of labour. This is not to say, of course, that more conventional ways to collect data did not exist; however, I would argue that they are significantly more labour-intensive than the method adopted in this study, and the extra workload is not compensated by better recall. One alternative way to start collecting

¹¹ The first query yielded 77 N(Deg)-ADJ compounds with 50 adjective types. The first iteration (based on the 50 adjective types) yielded 630 new compound types and 396 noun types. The second iteration (based on the new noun types) yielded 286 new compounds and 78 new adjective types. The third iteration (based on the 78 new adjective types) yielded 166 new compounds and 156 new noun types. The fourth iteration recalled 49 new compounds and 28 new adjective types. The fifth iteration recalled 22 new compounds and 14 new nouns. These 14 nouns did not give new results.

data would have been to use the N(Deg)-ADJ compounds discussed in Marchand (1960: 84–87) instead of the twenty subjectively selected nouns.¹² Marchand lists 28 compounds which can be considered to represent the N(Deg)-ADJ type. Of these 28 compounds, 24 were retrieved by my method, and the remaining four compounds are not included in COHA. We can therefore conclude that the recall of the method used in this study is at least on a par with the approach based on Marchand’s list of compounds. Furthermore, and importantly, it seems that the subjective selection of the first twenty nouns used in the corpus queries did not have a negative effect on recall.

To further assess the recall of my method, I examined yet another way to collect data by using frequency lists. First, I downloaded lists of 1,000 most frequent adjectives in COHA for two decades (the 1810s and the 1930s). I was interested in checking the data for the 1810s, because I was only able to retrieve five N(Deg)-ADJ types in the corpus for that period with my method (*snow-white*, *blood-red*, *clay-cold*, *sun-bright*, *silver-clear*). The 1930s, on the other hand, was chosen randomly to represent the twentieth century. The results were as follows: for the 1810s, my method recalled all the relevant instances of the N(Deg)-ADJ construction. For the 1930s, six compounds were not recalled (*God-awful*, *world-old*, *time-old*, *pillar-powerful*, *clam-silent*, *owl-solemn*). However, my method recalled seven compounds that were not found by the frequency list approach (*stone-deaf*, *porcelain-frail*, *apple-glossy*, *cream-puffy*, *harem-seductive*, *fox-sneaky*, *rock-sound*). Based on these comparisons, I would argue that my data collection method is not, in the very least, inferior to more conventional methods in terms of recall. It should be admitted, however, that the recall probably remains somewhat poorer for the PN(Deg)-ADJ constructs due to the fact that the proper noun

¹² I thank an anonymous reviewer who suggested that these compounds could be used to check the recall of the data collection method.

slot can in fact consist of two or more proper nouns (e.g. both the first and the last name of the paragon; i.e. a “proper name”), and therefore the likelihood of a PN(Deg)-ADJ construct being written with a hyphen is much lower than is the case with CN(Deg)-ADJ constructs (e.g. *pitch-black* vs. *Han Solo cool*). Because of this, I performed one last query where I used all the adjectives recalled in the earlier queries and searched the corpus for proper nouns followed by these adjectives using POS annotation.

A final issue that needs to be discussed in this context is the potential effect of corpus size on the results. In the diachronic case study discussed in section 4, the data are divided into ten periods, which are not perfectly equal in size. As it is well-known that the type frequencies of linguistic constructions do not increase linearly with corpus size (see e.g. Baayen 2009: 902–903), there are some points that I wish to make before proceeding with the analysis.

First, while the sub-corpora used in the case study are not perfectly balanced in terms of word counts, many of the periods are actually very well in balance. This is particularly true for the twentieth-century data, where the word counts fluctuate from 48.2 million words (1910–1929) to 48.9 million words (1970–1989) (see Table 2). Second, even if type frequencies do not increase linearly with corpus size, we do have a relatively good idea of what to expect. In their study of morphological productivity, Plag et al. (1999: 218) estimated that if a two-million-word corpus of written language includes c. 100,000 word types, a four-million word corpus would include c. 140,000 types. In order to double the type count of the original corpus, we would need a corpus five times as large as the original corpus (i.e. a ten-million-word corpus).¹³ As we shall

¹³ In Plag et al.’s article (1999), the relationship between corpus size and type frequency varied for each derivational affix studied, but the overall result remained the same: if you start with a sufficiently large

see in section 4, the changes in the productivity of the N(Deg)-ADJ construction are simply too substantial to be explained away by uneven corpus size. Having said that, I will study the data from different perspectives in order to show that all the evidence points to the same conclusion.

Table 2. Word counts of the sub-corpora used in the diachronic case study (COHA).

	1810– 1829	1830– 1849	1850– 1869	1870– 1889	1890– 1909	1910– 1929	1930– 1949	1950– 1969	1970– 1989	1990– 2009
Million words	8.1	29.8	33.6	39.5	43.7	48.3	48.6	48.3	48.9	57.4

4 THE N(DEG)-ADJ CONSTRUCTION IN THE CORPUS OF HISTORICAL AMERICAN ENGLISH

4.1 *Productivity of the CN(Deg)-ADJ construction in COHA*

Let us start our investigation by taking a look at the overall development of the CN(Deg)-ADJ construction in terms of frequency of use (Figure 2).

corpus, the type frequency will not double when the corpus size is doubled – the increase will be substantially smaller.

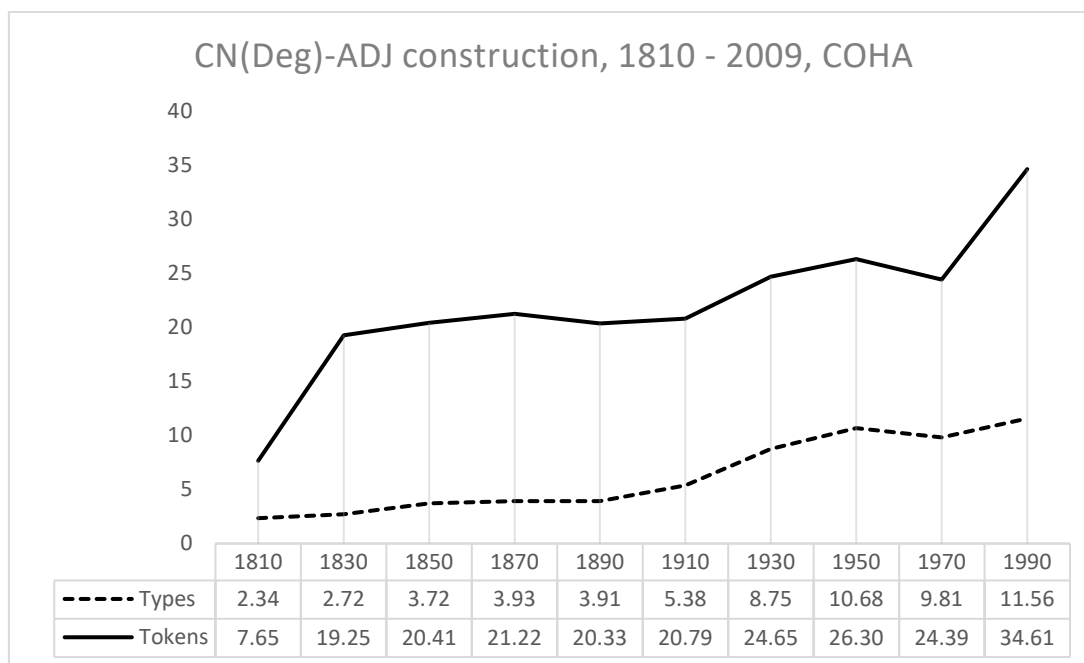


Figure 2. The type and token frequencies of the CN(Deg)-ADJ construction in 1810–2009, COHA. Normalised to 1,000,000 words.

Figure 2 shows a substantial increase in both the type and token frequencies of the construction from 1810 to 2009: the type frequency has increased from 2.34 (1810–1829) to 11.56 per one million words (1990–2009), while the token frequency has gone up from 7.65 tokens (1810–1829) to 34.61 (1990–2009). Overall, the nineteenth century represents a period of relative stability: if the first period is ignored,¹⁴ the token frequency remains stable throughout the century, and after an initial increase, the type frequency is also stable for the latter part of the century.

After a period of stagnation, both type and token frequencies begin to increase rapidly at the start of the twentieth century. The type frequency of the CN(Deg)-ADJ construction rises from 3.91 to 5.38 from 1890–1909 to 1910–1929, and further to 8.75

¹⁴ The early decades in COHA are not very well balanced, and it is a common experience in the corpus linguistic community that they often yield results that are not in line with a more general trend.

in 1930–1949. The token frequency of the construction likewise climbs from 20.79 in 1910–1929 to 24.65 in 1930–1949. As we are mainly interested in the productivity of the pattern, let us examine the data from the perspective of how many new CN(Deg)-ADJ compounds, new nouns, and new adjectives are attested in each period (i.e. compounds, nouns and adjectives that had not been attested in the previous periods).

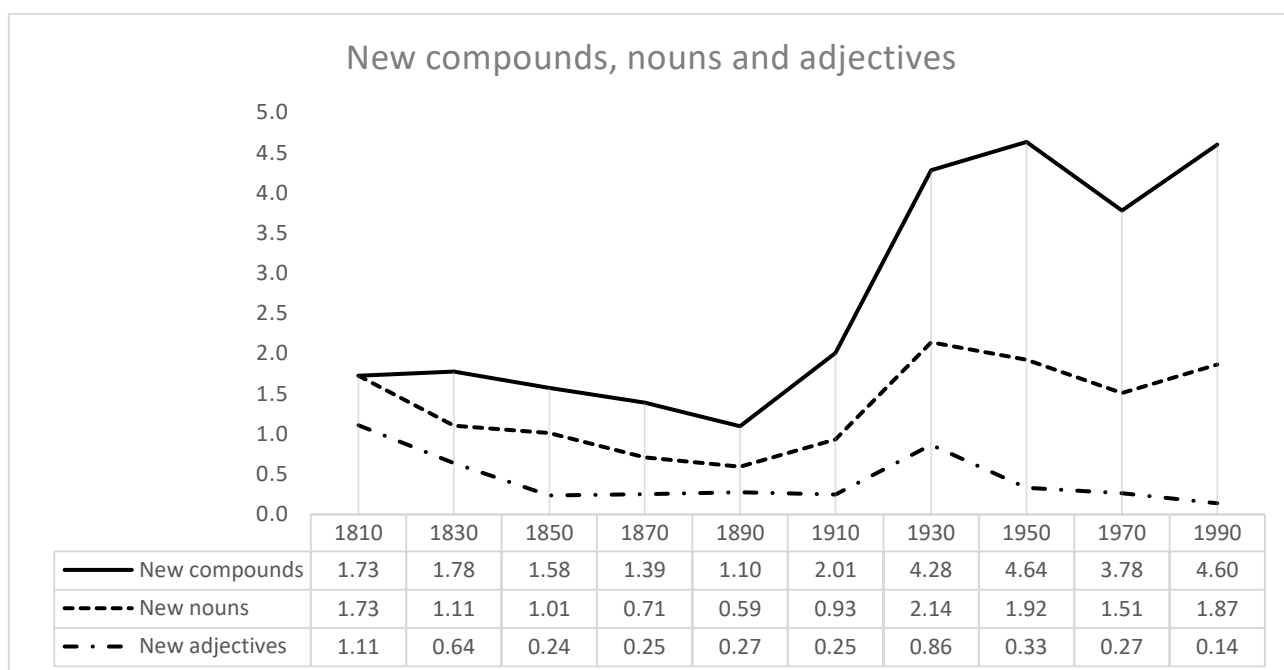


Figure 3. Productivity of the CN(Deg)-ADJ construction in terms of new compounds, nouns and adjectives in COHA. Normalised to 1,000,000 words.

We can see in Figure 3 that a major development indeed takes place in the late 1800s and the early 1900s. Most importantly, authors start to use a larger variety of nouns in a modifying function, and the number of previously unattested compounds also increases rapidly. The 1930s and 1940s represent a peak in the use of new nouns and adjectives, and the productivity of the pattern has remained high in terms of new compounds and new nouns until the present day.

Although it is not entirely clear how the development depicted in Figure 3 could be examined from a statistical perspective, the fact that the reference corpus constantly increases in size strongly suggests that we are witnessing a genuine increase in the productivity of the CN(Deg)-ADJ construction. If the frequencies remained stable across time, we would expect to see decreasing trends in all categories due to the increasing size of the reference corpus. This is indeed what we see in the nineteenth-century data, but the twentieth-century data show a contrary development with the type frequency of new compounds increasing from 1.10 in the 1890s to 4.28 in the 1930s.

As the developments depicted in Figures 1 and 2 are largely based on normalised type frequencies, the results can be questioned because the size of the sub-corpora varies somewhat (see the discussion in section 3). However, there are ways to complement the findings presented above. First, the entire corpus can be divided into two equally-sized sub-corpora. The first sub-corpus includes data from 1810 to 1929 (203.2 million words) and the second one data from 1930 to 2009 (203.0 million words). This way, we should be able to effectively eliminate the potential effect of any fluctuations in corpus size, although the precise dating of frequency changes naturally suffers in this approach. Table 3 summarises the findings from these two sub-corpora. The results are well in line with the data presented in Figures 2 and 3: both type and token frequencies show a substantial increase in the twentieth century.

Table 3. Type and token frequencies in two equally-sized sub-corpora in COHA.

	1810–1929	1930–2009
Type frequency	813	2,084
Token frequency	3,580	5,688

There are also other measures that can be used to gauge the productivity of the N(Deg)-ADJ construction. Baayen & Lieber (1991: 809) point out that productive word formation patterns typically have a high proportion of hapax legomena in corpus data (i.e. types that only occur once in the entire corpus), whereas unproductive patterns are characterised by a high number of high-frequency types (see also Baayen & Renouf 1996: 74). In other words, by studying the proportion of hapax legomena in each period, we can assess the changes in the productivity of the CN(Deg)-ADJ construction over time. Furthermore, by measuring productivity through hapaxes we are also able to mitigate the effect of fossilised types (e.g. *stone-cold*, *sky-blue*), which are not relevant to the question of increased productivity of the construction (see also Berg, forthcoming).

Figure 4 shows the increase in the proportion of CN(Deg)-ADJ hapaxes in COHA. Note that the proportion of hapaxes is measured for the entire corpus (from 1810 to 2009), not for each twenty-year period, and the measure should therefore not be as sensitive to the uneven word counts across the periods.

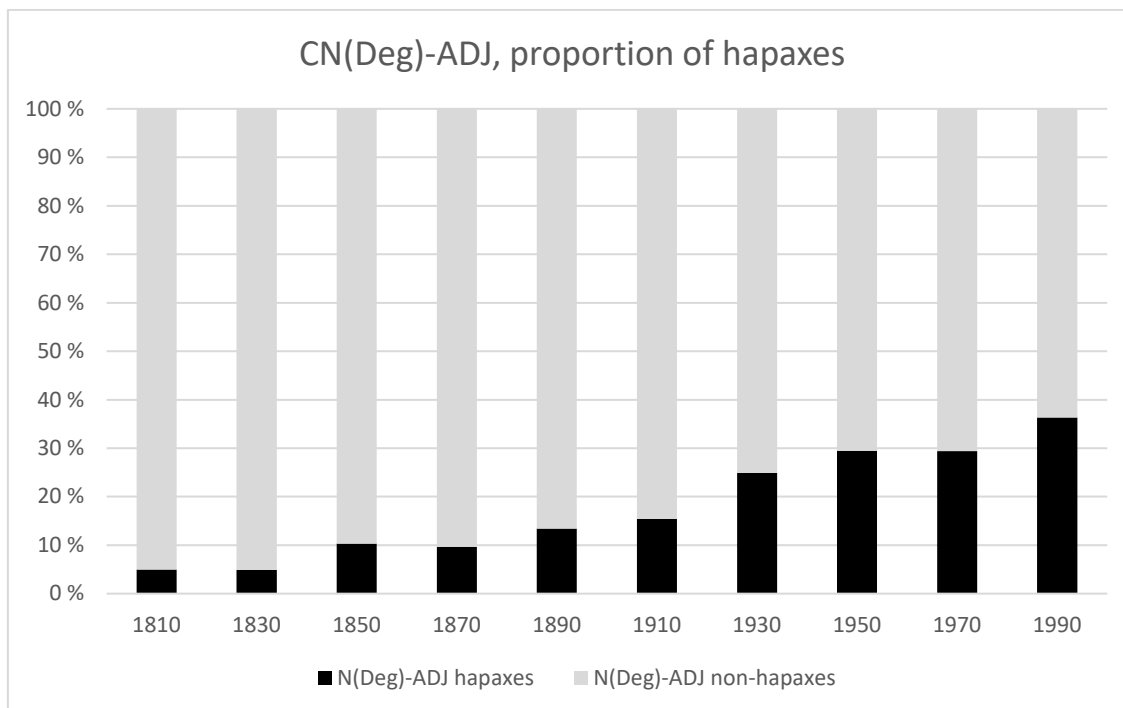


Figure 4. The proportion of CN(Deg)-ADJ hapax legomena in COHA.

The steady increase in the proportion of hapax legomena is clear evidence of the fact that the productivity of the CN(Deg)-ADJ construction has increased steadily for the period studied and that the increase has been particularly outstanding in the twentieth century. There is a sudden rise in the proportion of hapaxes from 1910–1929 (15.4%) to 1930–1949 (24.9%), and another one from 1970–1989 (29.4%) to 1990–2009 (36.3%).¹⁵ We will see in sections 4.2 and 4.3 that the first peak coincides with a semantic development that may have supported the emergence of the proper noun construction.

In addition to providing us with reliable data concerning the productivity of the CN(Deg)-ADJ construction, we can use the hapax data to see whether the increase in

¹⁵ Hapaxes in the two equally-sized sub-corpora can again be examined to complement the results. In 1810–1929, 11.8% of all CN(Deg)-ADJ types are hapaxes, while in 1930–2009 the proportion of hapaxes is 30.7%.

productivity is statistically significant by using a correlation test. Here, we first calculate the median of the type frequencies of the ten time periods in COHA, and then study the way in which the data are distributed for each period. A null hypothesis would be that there is no correlation between the ten time periods and type frequency: the frequencies should fluctuate randomly, that is, in some periods the number of hapaxes would be below the median while in others it would be above it; importantly, the variation would be unsystematic.

Table 4 shows the absolute frequencies of hapax legomena for the CN(Deg)-ADJ types in COHA. The trend is clear: the number of hapax legomena is below the median (31.5) in the first five periods and above it in the last five periods. The probability that the null hypothesis is true is extremely low ($p = 2/252 < 0.007937$),¹⁶ which provides robust evidence to the claim that the construction has become increasingly productive in the twentieth century. Indeed, the data suggest that the productivity of the construction may not even have peaked yet.

Table 4. CN(Deg)-ADJ construction. Number of hapaxes in COHA, 1810–2009. Absolute frequencies.

1810	1830	1850	1870	1890	1910	1930	1950	1970	1990
1	4	13	15	23	40	106	152	141	241

4.2 *Semantic change*

The increase in the productivity and usage of the CN(Deg)-ADJ construction has also resulted in constructional change in terms of the kinds of common nouns that are used

¹⁶ There are 252 ways to choose which five periods out of ten have a value above the median, and only two of the possible permutations are as extreme as what is observed. I thank Jukka Suomela for methodological assistance.

in the construction. In the earlier periods, the nouns typically denoted natural phenomena and inanimate entities (e.g. *sky-high*, *ice-cold*, *sun-bright*, *stone-cold*), as well as everyday objects and materials (e.g. *iron-strong*, *sheet-white*, *silk-soft*). In the late nineteenth century, and especially in the early twentieth century, human nouns start to be used in the construction. Examples (12) to (14) illustrate these new types, and Figure 5 shows the increase in the type frequency of human nouns in the corpus.

(12) Well I never cared to see a man **maiden-meek**. (COHA, Fiction, 1889)

(13) ... her **harem-soft and harem-seductive hand** with the tiny cream-colored wrist pressing the stocking-cap closer to the smooth black waves of hair.¹⁷
(COHA, Fiction, 1936)

(14) Don Juan put **his woman-soft hand** upon Ross's shoulder. (COHA, Fiction, 1950)

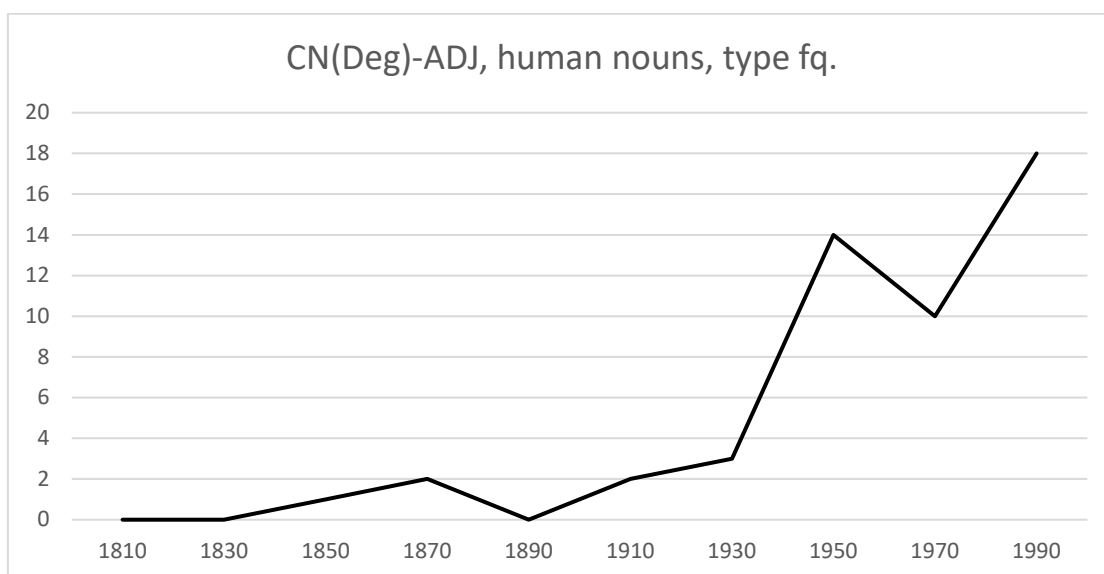


Figure 5. Human nouns used in the CN(Deg)-ADJ construction. Type frequency, 1810–2009, COHA.

¹⁷ *Harem-soft* and *harem-seductive* in example (13) are the only collective common nouns in the data.

It must be admitted that the frequency of human nouns is still very low (only 50 out of 586 noun types in the entire corpus are human nouns), but the data do show an increasing trend. There is one human noun that has become particularly productive since the 1950s: *baby* is found in the data in compounds like *baby-soft*, *baby-smooth* and *baby-bald*, while data from the most recent decades include examples like *addict-thin*, *celebrity-perfect*, *pageant-beautiful*, *refugee-skinny*, *model-thin*, *dancer-thin* and *Jedi-quick*. The increased frequency of human nouns leads us to the final topic of inquiry, the PN(Deg)-ADJ construction.

4.3 *The PN(Deg)-ADJ construction*

Compared to the CN(Deg)-ADJ construction, the PN(Deg)-ADJ construction is rare, and it has emerged much more recently. The first PN(Deg)-ADJ construct found in COHA is from 1856. Here, swell height is compared to Mount Olympus.

- (15) May the winds blow till they have wakened death! And let the laboring bark climb hills of seas **Olympus-high**, and duck again as low as hell's from heaven. (COHA, Fiction, 1856)

In addition to example (15), the corpus queries returned only one other token in the nineteenth-century data.

- (16) His **Croesus-bright scepter** has magical sway, Yester's indifference solicits to-day. (COHA, Fiction, 1881)

Interestingly, in both cases the proper noun refers to a mythical or a legendary referent, whom the authors assume to be familiar enough to the readers to be considered a paragon.¹⁸

Examples (15) and (16) also illustrate the two main types of proper nouns used in the PN(Deg)-ADJ construction: geographical nouns (*Olympus*) and human nouns (*Croesus*). There are only three PN(Deg)-ADJ constructs from the 1910s and the 1920s in the corpus, and in each case the noun slot is filled with a geographical noun (examples (17)–(19)). Although it is possible that geographical nouns may have initially been more common than human nouns in the PN(Deg)-ADJ construction (this would be in line with the results of Rosenbach’s (2007) study on noun modifiers), the numbers are unfortunately too small to draw any conclusions.¹⁹

(17) [...] an excruciating sweetness obtained only by the wallowing, walloping yellow-pink palm of a hand whose back was **Congo black** and shiny. (COHA, Fiction, 1914)

(18) It was hot? African hot, not **United States hot**? (COHA, Fiction, 1921)

(19) On the wet and dry issue, Mr. Smith appears to have drawn support from partisans on both sides. Down-State he had support of the Anti-Saloon League as **a Sahara dry**, while in Cook County the Crowe-Barrett Organisation, dripping wet, apparently gave him full support. (COHA, News, 1926)²⁰

Figure 5 shows that both human and geographical proper nouns have increased relatively steadily in the latter part of the twentieth century.

¹⁸ *Croesus-bright sceptre* represents a rare kind of usage. In the PN(Deg)-ADJ construction, the referent is typically directly associated with the property denoted by the adjective. In (16), by contrast, *Croesus* is only metonymically associated with brightness (through his association with gold).

¹⁹ Altogether 17 human nouns and 16 geographical nouns are used in the PN(Deg)-ADJ construction in the COHA data.

²⁰ In (17), the PN(Deg)-ADJ construct is used as a noun phrase. This is the only example of such usage in my data.

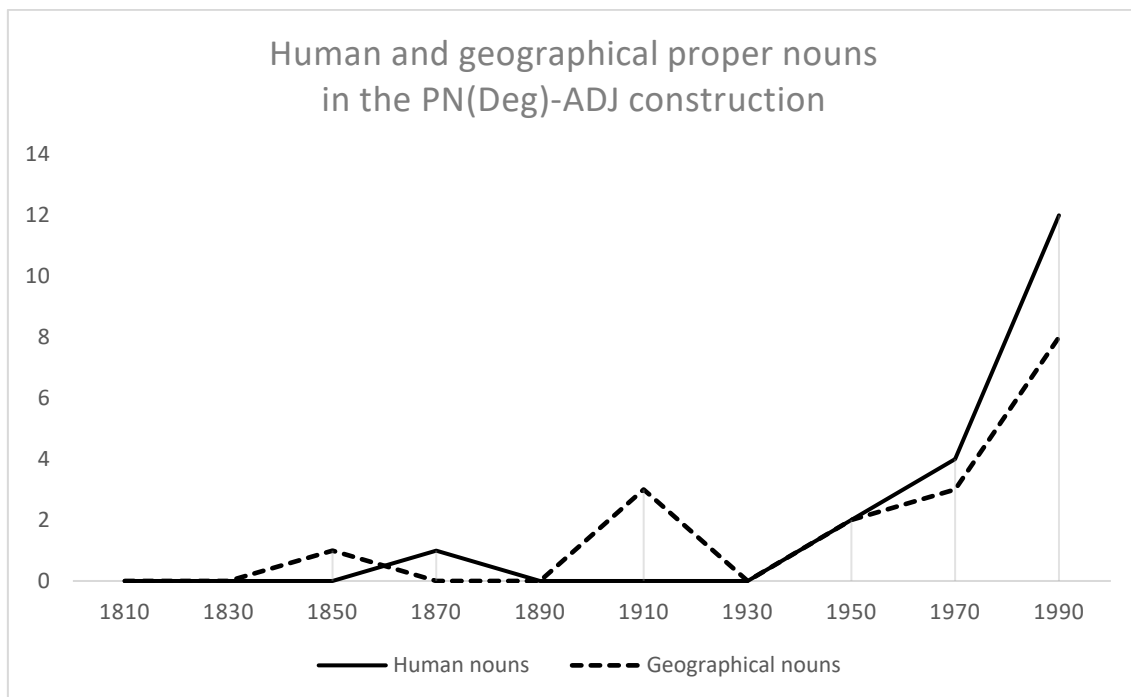


Figure 5. Human and geographical proper nouns in the PN(Deg)-ADJ construction. COHA. Absolute numbers.²¹

As can be seen in Figure 5, human nouns start to appear more regularly in the data since the 1950s. This is significant, as the frequency increase coincides with the period in which the frequency of human nouns peaks in the CN(Deg)-ADJ construction. Figure 7 shows that the frequency of human proper nouns closely follows the frequency of human common nouns that were depicted in Figure 5 above.

²¹ Human proper nouns in the last period include two collective nouns which refer to baseball teams (*Oakland bad, Detroit bad*). One token (*Nintendo fast*) is not included in the graph.

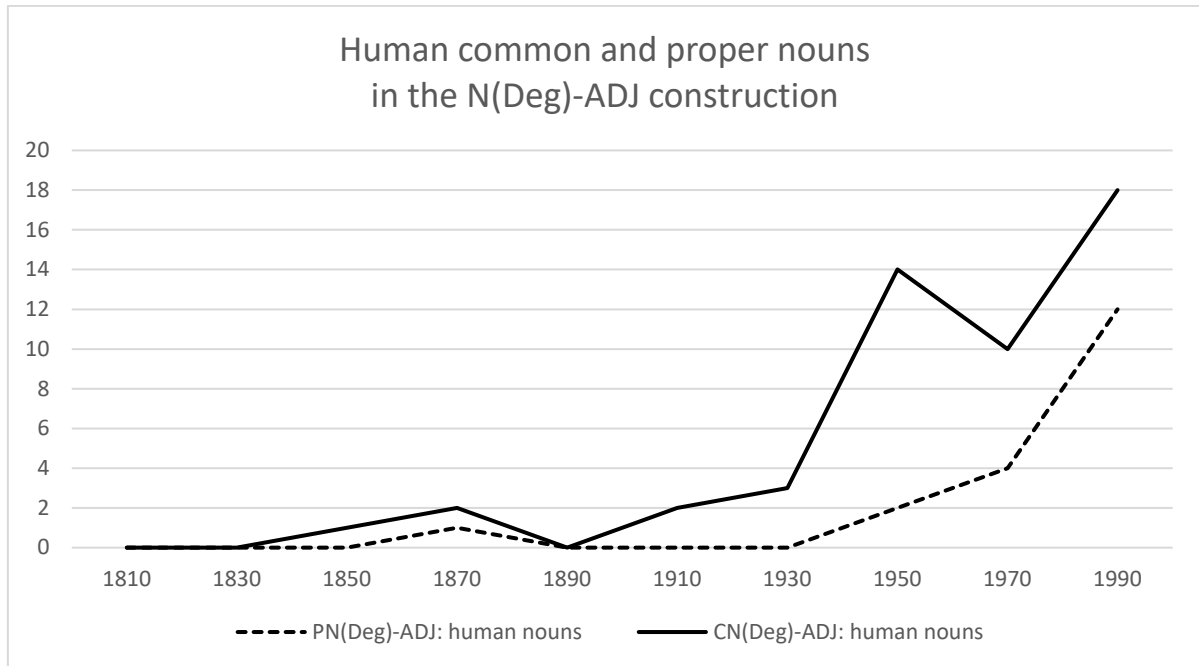


Figure 7. Human common and proper nouns in the N(Deg)-ADJ construction. COHA.

The increased frequency and gradual conventionalisation of the PN(Deg)-ADJ pattern has recently resulted in an interesting usage where there is a mismatch between the animacy value of the proper noun and the referent that is described by using the PN(Deg)-ADJ construct. In examples (20) to (25) we find inanimate objects evaluated with a PN(Deg)-ADJ compound that hosts an animate noun. In some cases, as in example (23), the author even plays with the different meanings of the adjective; here, *rich* is used to describe the richness of a ravioli dish by referring to the immense wealth of Oprah and Bill Gates. Although these uses are not common, they are reasonably well attested in different corpora and could be taken as further evidence of the conventionalisation of the PN(Deg)-ADJ construction.

- (20) All of a sudden, the most-isolated, least-sexiest football program in the Pac-12 became **Eva Longoria hot**. (GloWbE, US)

- (21) Yumzo. I love potato soup and this one's kinda making me drool all over the keyboard. Maureen, I want the recipe:) Oh and whats up with rose bushes growing **Heidi Klum tall?** (GloWbE, CAN)
- (22) The new players (well, except one) made the difference and helped us win in a **Susan Boyle ugly way**.²² (GloWbE, GB)
- (23) The ravioli (\$24) is rich. Not just **Oprah rich, Bill Gates rich**. The sauce is heavily creamy [...] (iWeb, CAN)
- (24) Ever open up a G5 tower or mac pro? It's prettier inside then [sic] it is outside. And not just **Penelope Cruz pretty**, but pretty in terms of the aesthetics of cable runs, expansion bays, etc. The beauty is in the excellence of execution. (GloWbE, US)
- (25) If you're tired of riding, Stovepipe offers **an Einstein-smart option**. (Gary McKechnie: Great American Motorcycle Tours, p. 370)

Another way to measure the conventionalisation of a construction is to study its global diffusion. So far, most of the examples of the PN(Deg)-ADJ pattern are from American English, and this raises the question of whether we are dealing with a usage that is idiosyncratic to AmE.

To examine this question, I carried out a small-scale pilot study of the PN(Deg)-ADJ construction on the GloWbE corpus. As GloWbE is five times larger than COHA, it was not possible to collect data by using the method described in section 3. Furthermore, as paragons are culturally defined, using proper nouns as the basis of corpus queries would have introduced bias. Consequently, the corpus search was based

²² An anonymous reviewer wondered if attributive uses like this are a recent development. The data are too scarce to answer this question conclusively, but as all the PN(Deg)-ADJ constructs in COHA appear in predication, this is a sensible hypothesis. If confirmed, examples like (22) and (25) could be taken as further indication of conventionalisation.

on a set of forty adjectives, which were chosen based on their relatively frequent use in the PN(Deg)-ADJ construction in previous corpus queries. The adjectives included in the query were: *amazing, athletic, bad, beautiful, big, bright, chic, cool, crazy, cute, dumb, edgy, elegant, evil, famous, fast, fat, funny, good, gorgeous, handsome, hot, huge, intense, nasty, pretty, quick, rich, scary, sexy, skinny, slow, strange, stupid, tall, thin, tough, ugly* and *wonderful*.

In total, twenty-eight of the forty adjectives were used in the PN(Deg)-ADJ construction in the GloWbE data (a total of 74 tokens). Interestingly, however, the construction was only found in ten of the twenty varieties included in the corpus. The results are presented in Table 5 (the varieties with no attested tokens are left out).²³

Table 5. The PN(Deg)-ADJ construction in the global varieties of English (GloWbE) based on a selection of twenty adjectives.

	US	CA	GB	IE	AU	NZ	SG	PH	HK	NG
N	30	9	16	5	4	2	2	1	2	3
per 1,000,000	0.08	0.07	0.04	0.05	0.03	0.02	0.05	0.02	0.05	0.07

The results in Table 5 are of course very preliminary, but based on them, it could be tentatively argued that the construction is particularly frequent in North America. What is most interesting, though, is the fact that the construction could not be found at all in Indian, Sri Lankan, Pakistani, Bangladeshi, Malaysian, South African, Ghanaian, Kenyan, Tanzanian or Jamaican English, even though the combined word count for these varieties exceeds that of American English in the corpus (c. 438 million words vs.

²³ The abbreviations stand for the United States, Canada, Great Britain, Ireland, Australia, New Zealand, Singapore, Philippines, Hong Kong and Nigeria, respectively.

c. 386 million words). The data therefore suggest that the degree of conventionalisation of the PN(Deg)-ADJ construction varies both within the “core varieties” of English and across the global varieties. Interestingly, most, though not all, of the examples attested in the global varieties refer to globally famous people instead of locally available paragons, as in (26) and (27), which are taken from the Singapore and Hong Kong sections of GloWbE, respectively.

(26) I don't wanna see him go **Britney S. crazy** over these dumb people. (GloWbE, SG)

(27) But he is **Daniel Day Lewis good**, he is **Sean Penn good**. (GloWbE, HK)

Examples (28) and (29), on the other hand, feature paragons whose role is negotiated in a more local discourse context. Example (28) is taken from a blog post published in an Irish sports website, which allows the writer to use a former cricket player and TV pundit, *Bill Lawry*, as a paragon. In (29), also from Ireland, the writer compares the playing style of two Irish fiddle players, taking *Martin Hayes* as an exemplar for slow playing. The currency of paragons like these is heavily context-dependent, and they are indeed much less frequently used in the construction than, say, world-renowned actors and musicians that were exemplified in (26) and (27).

(28) Man you have a big nose – not quite **Bill Lawry big**, but still pretty big. (GloWbE, IE)

(29) Listen to a range of Kevin Burke's jig playing – some nice slowish (not '**Martin Hayes slow**' though) jigs with excellent quality (e.g the minor version of The Rambling Pitchfork). (GloWbE, IE)

5 DISCUSSION AND CONCLUSIONS

In this paper I have examined the recent history of the N(Deg)-ADJ construction, investigating the individual developments of two patterns where the noun slot is filled either by a common noun or a proper noun. The results show that the type and token frequencies of the CN(Deg)-ADJ construction have increased substantially in the twentieth century. These results, together with a large increase in the proportion of hapax legomena, are evidence of growing productivity; a trend that starts at the beginning of the twentieth century and continues to the present day. This increased productivity is all the more interesting considering that while the construction has been in existence since the Old English period, both the results of this study and comments made in previous literature show that it is only now that language users have started to explore its full potential by creating a multitude of new forms and using a more varied range of nouns to express degree. It must be admitted, however, that earlier periods of English have not been systematically studied from a corpus linguistic perspective, and our understanding of the productivity of the construction may therefore change through future studies.

The emergence of the PN(Deg)-ADJ pattern can be regarded as a consequence of a micro-development affecting the N(Deg)-ADJ construction, on the one hand, and of two macro-trends affecting English grammar, on the other. The micro-development concerns the introduction of human nouns into the CN(Deg)-ADJ construction, a process which started in the early twentieth century (e.g. *maiden-meek*, *woman-soft*). I would argue that it is logical to assume that this change also facilitated the use of human proper nouns in the construction. The macro-trends supporting the emergence of the PN(Deg)-ADJ pattern have to do with changes affecting the use of proper nouns in

English as well as changes in the frequency of word formation patterns. First, Rosenbach (2007: 163) shows that human noun modifiers (both common and proper nouns) of the type *rebel army* or *Bush administration* have steadily become more frequent since the start of the twentieth century. Second, there is evidence that compounding has in general become more frequent in English. For example, Wald & Besserman (2002) show that the productivity of VV-compounds (e.g. *spell-check*, *hang-glide*) has increased in the twentieth century. Günther (forthcoming), on the other hand, shows that the increase in the frequency of premodifying phrasal compounds (e.g. *important-but-hard-to-remember details*) starts in the early twentieth century. These developments coincide almost perfectly with the frequency increase of the N(Deg)-ADJ construction examined in this paper, and it seems unlikely that they should be totally unrelated.

One question that was explored in this paper is the status of the PN(Deg)-ADJ pattern as a construction, and a key point in this discussion was the degree to which the pattern has become conventionalised. In section 2, I cited speaker judgments and humorous uses in books and television as evidence of conventionalisation, but there are also some complicating factors, which I believe can be explained by the fact that the pattern has developed only recently. First, there is clearly some variation with regard to native speaker judgments. Some speakers report that although constructs like *Bill Gates rich* are arguably more expressive than, say, *extremely rich*, there is nothing particularly special about them. Other speakers maintain, however, that they find such usage acceptable only in restricted contexts where the speaker tries to be particularly creative or clever. Clearly, the pattern is better entrenched for some speakers than for others, and those who find PN(Deg)-ADJ compounds to be perfectly acceptable can also use it in

new and innovative ways. For instance, describing rose bushes as *Heidi Klum tall* or computer hardware as *Penelope Cruz pretty* requires that the constructional degree meaning is foregrounded and the concrete associations between the paragon and its referent backgrounded. Although such usage could also be analysed in terms of loosening of pragmatic constraints, it might as well be interpreted as advanced entrenchment.

The second complicating factor concerning the conventionalisation and constructional status of the PN(Deg)-ADJ pattern arises from the pilot study on the global varieties of English. Here, the data suggest three things: i) the pattern is particularly well attested in the North American varieties of English, ii) it is attested to a somewhat lesser degree in the other “core” varieties, and iii) it is not attested at all in ten of the fourteen varieties of English that are spoken in different parts of Asia and Africa. Considering that we are dealing with a pattern whose development only started approximately one hundred years ago, this partial global diffusion is probably something to be expected. These results also underscore the fact that whatever weight is given to speaker judgments in questions related to the meaning and use of a construction, variation both within and between different communities of speakers is expected.

Even though the conventionalisation – and constructionalisation – of the PN(Deg)-ADJ pattern may still be ongoing, we can already see many parallels in the way it is used and the use of intensifiers of a more regular kind. Tagliamonte (2008: 362–363) provides a summary of some socio-cultural correlates that have been particularly associated with the use of intensifiers in the literature, including gender differences, age-related usage, colloquial/non-standard varieties, emotive usage and in-

group membership. There certainly seems to be a colloquial flair to PN(Deg)-ADJ compounds. This shows, for instance, in the fact that they are not attested in the non-fiction and newspaper registers in COHA. Furthermore, as was already discussed in section 2, the pragmatic presuppositions associated with the PN(Deg)-ADJ compounds make the construction useful for signalling speaker/hearer alignment and in-group/outgroup membership. For instance, calling someone *Sarah Palin stupid* is a clear indication of political alignment, while assessing someone or something to be *Lady Gaga strange* can place the speaker in a specific position with regard to, say, people following fashion. Such socio-cultural correlates would be interesting to study in future research. Furthermore, as we are clearly witnessing ongoing language change, it would be interesting to revisit the construction in the not-too-distant future to see if its productivity keeps increasing and if it continues to spread to the varieties of English in which it was not attested in this study.

Turo Vartiainen

Department of Languages

University of Helsinki

PO Box 3

00014, University of Helsinki

Finland

E-mail: turo.vartiainen@helsinki.fi

REFERENCES

- Antonopoulou, Eleni. 2004. Humor theory and translation research: Proper names in humorous discourse. *Humor* 17(3), 219–255.
- Baayen, R. H. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics: An International Handbook*, 899–919. Berlin: Mouton de Gruyter.
- Baayen, R. H. & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29(5), 801–843.
- Baayen, R. H. & Antoinette Renouf. 1996. Chronicling the Times: Productive lexical innovations in an English newspaper. *Language* 72(1), 69–96.
- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge Studies in Linguistics 155. Cambridge: Cambridge University Press.
- Berg, Kristian. Forthcoming. Productivity, vocabulary size, and new words. A response to Säily (2016). *Corpus Linguistics and Linguistic Theory*.
- Bergen, Benjamin & Kim Binsted. 2004. The cognitive linguistics of scalar humor. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture and mind*, 79–92. Stanford, CA: CSLI.
- Breban, Tine. 2018. Proper names used as modifiers: A comprehensive functional analysis. *English Language and Linguistics* 22(3), 381–401.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

- Cappelle, Bert. 2017. What's pragmatics doing outside constructions? In Ilse Depraetere & Raphael Salkie (eds.), *Semantics and pragmatics: Drawing a line*, 115–151. Amsterdam: Springer.
- Chapman, Don & Ryan Christensen. 2007. Noun-adjective compounds as a poetic type in Old English. *English Studies* 88(4), 447–464.
- Claridge, Claudia. 2011. *Hyperbole in English: A corpus-based study of exaggeration*. Cambridge: Cambridge University Press.
- Coulson, Seana. 2001. *Semantic leaps. Frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.
- Cowie, Claire & Christiane Dalton-Puffer. 2002. Diachronic word-formation and studying changes in productivity over time: Theoretical and methodological considerations. In Javier E. Díaz Vera (ed.), *A changing world of words: Studies in English historical lexicography, lexicology and semantics*, 410–437. Amsterdam: Rodopi.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Davies, Mark. (2008–) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. <https://corpus.byu.edu/coca/>.
- Davies, Mark. (2010–). *The Corpus of Historical American English: 400 million words, 1810-2009*. <https://corpus.byu.edu/coha/>.

- Davies, Mark. 2013b. *Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries (GloWbE)*. <https://corpus.byu.edu/glowbe/>.
- Davies, Mark. (2018-) *The 14 Billion Word iWeb Corpus*. <https://corpus.byu.edu/iWeb/>.
- Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6(3), 353–375.
- Fogelin, Robert J. 1988. *Figuratively speaking*. New Haven: Yale University Press.
- Glucksberg, Sam. 2001. *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press.
- Glucksberg, Sam & Boaz Keysar. 1990. Understanding metaphorical comparisons: Beyond similarity. *Psychological Review* 97(1), 3–18.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7(5), 219–224.
- Goldvarg, Yevgeniya & Sam Glucksberg. 1998. Conceptual combinations: The role of similarity. *Metaphor and Symbol* 13(4), 243–255.
- Günther, Christine. Forthcoming. A difficult to explain phenomenon: Increasing complexity in the prenominal position. *English Language and Linguistics*.
- Günther, Christine, Sven Kotowski & Ingo Plag. Forthcoming. Phrasal compounds can have adjectival heads: Evidence from English. *English Language and Linguistics*.

- Hilpert, Martin & Holger Diessel. 2017. Entrenchment in Construction Grammar. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning. How we reorganize and adapt linguistic knowledge*. Washington, DC: De Gruyter.
- Israel, Michael, Jennifer Riddle Harding & Vera Tobin. 2004. On simile. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture and mind*, 123–135. Stanford, CA: CSLI.
- Kay, Paul. 2013. The limits of (Construction) grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 32–48. Oxford: Oxford University Press.
- Lakoff, George. 1987. *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago (IL): University of Chicago Press.
- Lambrecht, Knud. 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.
- Lauwers, Peter & Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49(6), 1219–1235.
- Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lipka, Leonhard. 1966. *Die Wortbildungstypen WATERPROOF und GRASS-GREEN und ihre Entsprechungen im Deutschen*. Tübingen: University of Tübingen.

- Marchand, Hans. 1960. *The categories and types of Present-day English word-formation: A synchronic-diachronic approach*. Wiesbaden: Otto Harrassowitz.
- Michaelis, Laura A. 2002. Headless constructions and coercion by construction. In Elaine J. Francis & Laura A. Michaelis (eds.), *Mismatch: Form-function incongruity and the architecture of grammar*, 259–312. Stanford (CA): CSLI Publications.
- Norrick, Neal R. 2010. Pear-shaped and pint-sized: Comparative compounds, similes and truth. In Armin Burkhardt & Brigitte Nerlich (eds.), *Tropical truth(s). The epistemology of metaphor and other tropes*, 213–226. Berlin: De Gruyter.
- Ortony, Andrew. 1979. Beyond literal similarity. *Psychological Review* 86(3), 161–180.
- Paradis, Carita. 2000. It's well weird. Degree modifiers of adjectives revisited: The nineties. In John Kirk (ed.), *Corpora galore: Analyses and techniques in describing English*, 147–160. Amsterdam: Rodopi.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Plag, Ingo, Christiane Dalton-Puffer & Harald Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2): 209–228.
- Rosenbach, Anette. 2007. Emerging variation: Determiner genitives and noun modifiers in English. *English Language and Linguistics* 11(1), 143–189.

- Säily, Tanja. 2014. *Sociolinguistic variation in English derivational productivity: Studies and methods in diachronic corpus linguistics*. Mémoires de la Société Néophilologique de Helsinki XCIV. Helsinki: Société Néophilologique.
- Searle, John R., 1993. Metaphor. In Andrew Ortony (ed.), *Metaphor and Thought*, 83–111. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12(2), 361–394.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Trousdale, Graeme. 2013. Multiple inheritance and constructional change. *Studies in Language* 37(3), 491–514.
- Tversky, Amos. 1977. Features of similarity. *Psychological Review* 84(4), 327–352.
- Wald, Benji & Lawrence Besserman. 2002. The emergence of the verb-verb compound in twentieth century English and twentieth century linguistics. In Donka Minkova & Robert P. Stockwell (eds.), *Studies in the history of the English language: A millennial perspective*, 417–447. Berlin: Mouton de Gruyter.