LEGE ARTIS

Language yesterday, today, tomorrow

Vol. VII. No 1 2022

SYNTHETIC AND ANALYTIC ADJECTIVE NEGATION IN ENGLISH SCIENTIFIC JOURNAL ARTICLES: A DIACHRONIC PERSPECTIVE[1]

Katrin Menzel*, Marie-Pauline Krielke, Stefania Degaetano-Ortlieb

Saarland University, Saarbrücken, Germany

Corresponding author*

Bibliographic description: Menzel, K., Krielke, M.-P. & Degaetano-Ortlieb, S. (2022). Synthetic and analytic adjective negation in English scientific journal articles: A diachronic perspective. In *Lege artis. Language yesterday, today, tomorrow. The journal of University of SS Cyril and Methodius in Trnava*. Trnava: University of SS Cyril and Methodius in Trnava, 2022, VII (1), p. 157-213. ISSN 2453-8035 DOI: https://doi.org/10.34135/lartis.22.7.1.05

Abstract: This paper addresses the development of synthetic and analytic adjective negation in a corpus of English scientific articles from the mid-17th century towards the end of the 20th century. Analytic patterns of adjective negation are found to become less frequent in the language of scientific articles, but more conventionalised in their textual contexts. Conversely, prefixed negated adjectives are identified as more frequent and more diverse with regard to their contexts.

Key words: adjective negation, prefixation, clausal negation, scientific English, corpus-based diachronic analysis, surprisal.

1. Introduction

In this article, we explore the changes in the use of English adjective negation patterns with affixal and non-affixal negation markers. We specifically focus on synthetically negated adjectives with the prefix *un-* or *non-* (e.g., *unavoidable*, *non-magnetic*) and analytic patterns of adjectives after the negation marker *not* (e.g., *not avoidable*, *not*

[1] The work for this article was performed in connection with a multi-phase research project funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102 "Information Density and Linguistic Encoding". The authors are primarily involved in its subproject B1: "Information Density and Scientific Literacy in English – Synchronic and Diachronic Perspectives" with a focus on linguistic densification in the evolution of scientific writing in English from the 17th century to the present. The authors also wish to thank an anonymous reviewer whose constructive questions and detailed comments on an earlier draft of the manuscript led to several clarifications.

magnetic) in a diachronic corpus of scientific journal articles. Bauer et al. (2013:4) acknowledge that specialised languages of scientific disciplines do fall under the scope of word formation research, but they deliberately exclude the discussion of formatives that are productive in highly scientific or technical fields from their reference volume on contemporary English morphology. One of the reasons for this might be the limited availability of sufficient specialised corpus data for analyses. Negation and word formation in general, including affixal negation and negative affixes in specialised languages, continue to be presented in the literature as marginal topics illustrated only with a few examples.

We would like to address this research gap by putting the focus on the above-mentioned synthetic and analytic adjective negation patterns in English scientific articles as a particular type of scholarly publications that have evolved considerably over the course of time. Our general question within the context of the project "Information Density and Scientific Literacy in English –Synchronic and Diachronic Perspectives" to which this case study is related is whether English scientific writing in journal articles has become more informationally dense over time. Academic authors and readers of scientific articles can be expected to have shared specialised and register-specific knowledge including knowledge of the previous discourse. Choosing shorter, more condensed, and less explicit synthetic forms, when there is an option to do so, should therefore contribute to higher economy of expression in scientific writing. Authors of academic articles may use such register-specific strategies to manipulate the information density distribution of selected linguistic structures with regard to the expectations and processing capacity of their addressees. We therefore assume that changes in the use of negation variants towards more synthetic patterns contribute to a trend for informational densification in English scientific writing. Synthetic forms of adjective negation as shorter forms of encoding will show higher information density than adjectives negated by *not*. We expect to find more of synthetic adjective negation in more condensed patterns and a smaller number of less dense analytic adjective

negation patterns.

For the purposes of this article we restricted our analysis to adjectives negated with *un-* or *non-* as these negation markers have been identified as the two most common negative affixes in adjectives (cf. Section 2.5). To our knowledge, there are no previous studies that focus on the usage contexts of these two affixes in particular. *Un-* has always been the most frequently used negative prefix in English. Already in Old English the number of lexemes in *un-* was very large, and *un-* remained one of the major formative elements in English (cf. OED entry on *un-*, prefix$_1$). *Non-* has increasingly gained in productivity and has become an equally important negation marker in Present-Day English (PDE) that is also classified as one of the major formative elements in English (cf. OED entry on *non-*). In terms of affixal or non-affixal negation, variants of negated adjectives are not always entirely semantically equivalent (cf. the example of *not likely* and *unlikely* in 2.3 or Tottie's (1999) discussion of *not happy* and *unhappy* emphasising that the scalarity of some predicates can make certain variants semantically non-equivalent). Nevertheless, corpus-linguistic, diachronic research into both types of negation is a viable undertaking that will lead to a better understanding of the constraints that may influence the variation between the two types (ibid., 233). The usage of phrasal elements such as non-clausal noun modifiers (e.g., adjectives in noun phrases) is typically associated with structural complexity (Biber & Gray 2010: 6). The development of patterns with analytic negation markers vs. more condensed ones with synthetic negation markers in English scientific journal articles is interesting with regard to their contribution to information density and to the relationship between word-internal, phrasal, and clausal complexity.

Tottie's hypothesis (1999: 234) was that variation between affixal and non-affixal negation of adjectives would over time reflect the existence of two stable systems in almost complementary distribution as scope, stylistic, syntactic, and lexico-semantic constraints continuously dictate the choice between such variants. We assume that

 ISSN 2453-8035

these linguistic forms are also affected by register-specific contextual probabilities and expect to see a trend towards an increased use of affixal negation in the language of academic journals. Adjectives negated with prefixes contribute to achieving economy in terms of the number of words that are used, i.e. a shorter code is used. This may also result in higher information locality compared to *not* + adjective, where *not* might be further away from the adjective and therefore possibly overlooked. We assume that cognitive constraints on information flow play a role in forming conventionalised, syntactically condensed patterns of negated adjective constructions adopted by a specialised scientific discourse community over time.

Functional motivations probably gave rise to a number of specialised uses of synthetically negated adjectives. In scientific research articles, adjectives with negative prefixes such as *un-* or *non-* in attributive positions are an important means to increase word-internal and phrase-internal compression of semantic information. This is done in order to achieve a higher degree of both word-internal and phrasal complexity and to reduce clausal complexity of alternative clausal constructions with the negation marker *not* in terms of sentence length, subordination structures, and the use of non-affirmative clause types (e.g., *"non-magnetic metals", "unavoidable difficulties"* vs. *"metals which are not magnetic", "difficulties are not avoidable"*). In this article, we would like to draw special attention to densification strategies within adjective constructions themselves and on strategies related to the word-internal, morphological level in connection with negation.

Especially Late Modern English (LModE) has been described as a period where a remarkable trend of densification via a changing usage of premodifying and postmodifying structures (e.g., higher frequencies of nominal premodifiers that could alternatively have been expressed by larger linguistic units such as relative clauses) started in registers such as newspapers and academic prose and subsequently spread to other linguistic contexts (Smitterberg 2021: 188). One of the consequences of such

densification processes that we expect to see in our corpus data already from the end of the Early Modern English period (EModE) onwards is that scientific language increasingly employed numerous negated adjectives as modifiers of head nouns. In this way, negated information has become generally more compressed into complex phrasal constructions enabling a tighter information structure (cf. Biber & Gray 2016: 4), while clausal negation patterns probably became restricted to fewer contexts. The usage of multimorphemic premodifiers such as synthetically negated adjectives as attributes in noun phrases, which often replace less compressed variants of clausal constructions, has an influence on the predictability of the upcoming words and on processing effort. Here we specifically consider **surprisal** (cf. Section 4.1) as an indicator of the predictability of words in context. Basically, higher predictability of a word in a given context leads to a lower surprisal score of this word, while lower predictability of a word in a given context means higher surprisal. As surprisal has been shown to be proportional to the cognitive effort of processing a word (cf. Hale 2001; Levy 2008), we can state that higher surprisal is related to higher cognitive effort.

The corpus used for the present article provides relevant data for exploring negated adjective constructions from late EModE to PDE in scientific articles. The **Royal Society Corpus** (RSC 6.0.1, cf. Fischer et al. 2020; Kermes et al. 2016; Menzel et al. 2021) consists of digitised texts from scientific journals such as the "Philosophical transactions" and the "Proceedings" of the Royal Society of London from 1665 to 1996. The authors are from the United Kingdom, the Commonwealth and beyond. We combine frequency-based analyses and information-theoretic modelling.

The structure of this article is as follows: Section 2 gives an overview of relevant background information from the literature on negation via affixes and other negation markers in English and discusses previous studies on the selected negation phenomena. In Section 3, we set out our hypotheses on the linguistic effects of the usage of affixal and non-affixal adjective negation in scientific writing. We describe our methodology

 ISSN 2453-8035

and data set in Section 4. The results of the analyses are given in Section 5. Section 6 contains our conclusions and an outlook for the future research.

2. Adjective negation in English

This section will give a broad outline of publications on synthetically negated adjectives, individual negative affixes, and non-affixal, analytical negation to establish a theoretical framework for our analysis.[1] We will also present an overview of previous psycholinguistic and corpus studies on negation in English.

2.1 Synthetically negated adjectives

One important type of lexical negation in English, i.e. negation of individual words, is synthetic (morphological) negation. Apart from prefixation, suffixation, and compounding, Cartoni & Lefer (2011) also mention conversion (e.g., *to dust = to remove dust*), but prefixation is the most common type of morphological negation in English (Joshi 2020: 80). Many sources on negation in English address primarily analytic forms such as clausal negation with *not* from a philosophical and/or linguistic perspective, but not particularly the negation of specific clause, phrase, or word types, e.g., those that involve adjectives. Fewer scholars address synthetic negation (cf., for instance, Hulse 2010: 42-44 for a selection of references). Some sources discuss diachronic aspects of negation (e.g., Croft 1991; Horn 2001; Mazzon 2004). Various sources on antonymy patterns also discuss the possibility of creating antonymy through negation and mention prefixation as an important process for the formation of morphologically related antonyms of adjectives (e.g., Charles & Miller 1989; Jones 2003; Murphy & Andrew 1993). Standard grammars and various linguistic studies list different means of negation, but not many works address the relationship between the different types and the question of when they may function as semantically equivalent variants.

 ISSN 2453-8035

English has various prefixes that can attach to adjectives and other word classes with different degrees of productivity to mark negation in general and specialised language, e.g., *a(n)-*, *de-*, *dis-*, *in-* (and its variants *il-*, *im-*, *ir-*), *non-*, *un-*$_1$, *un-*$_2$ (cf. Dixon 2014: 71; Marchand 1969; Zimmer 1964). Hulse (2010: 77) points out that some sources also list forms such as *anti-*, *contra-*, *counter-*, *dys-*, *ex-*, *extra-*, *mal-*, *mis-*, *para-*, and *sub-* as potential negative prefixes, and one might even add the rarely used *e-* (as in *ecaudate*). They all express negation to a certain degree, but in contrast to a free negation marker such as *not*, they potentially carry additional semantic characteristics to express nuances and different aspects in addition to negation, e.g., evaluative judgements. There is also the adjective-deriving suffix *-less* and the form *-free* (which is in between a suffix and a compound-forming element and also sometimes referred to as a combining form or a suffixoid, cf. Lieber 2004: 109) that can mark negation in adjectives synthetically and create English adjectives from nouns (e.g., *guiltless*, *carefree*) and thus change the word class of the base (in contrast to negative prefixes). Negated adjectives generally have several potential meanings. For the most part they foreground diverse denotational aspects of negation that are determined both by their internal elements and their respective context, e.g., their collocations.

Within this article, we can offer only a brief commentary on negation of adjectives with lexical antonyms (e.g., *hot* ≠ *cold; alive* ≠ *dead*), that is on a different means of creating opposites of adjectives, other than the formation of synthetically negated adjectives. The formation of synthetically negated adjectives that are potential words generable by regular word formation processes may in certain cases be "blocked" if lexical, morphologically unrelated antonyms of the non-negated form of the adjectives exist as rival forms. For instance, **unabsent/inabsent/non-absent* is probably non-existent not only due to the prefix *ab-*, but also due to the existence of *present* as an exact antonym of *absent* (cf. Kjellmer 2005: 160). *Absent* can only be negated analytically with *not* or by choosing its lexical antonym. From a PDE perspective, a few cases have become etymological cases of negation in which negation is marked rather implicitly and the

 ISSN 2453-8035

negation marker and morpheme boundary are not fully transparent any longer (e.g., *disgusting*). Some negated forms of adjectives may drop out of usage (e.g., *ungood* meaning 'bad, evil', has been used from Old English onwards, but has become rare in PDE, cf. its OED entry) or are chosen primarily for stylistic purposes and as literary devices (e.g., OED quotations containing *ungood* in modern use come from literary contexts as in Orwell's [1954] fictional language "Newspeak", which also includes the coinage *uncold (= warm)* that has no OED entry). Hulse (2010: 27) quotes a usage of *undrunk* that has not become an established word in English with the opposite meaning of "intoxicated" but was easily understood by the audience in a specific context.

A negated adjective can have variants with different prefixes ("morphological doublets" (Okada 2010: 349) or "multiple negative derivatives" (Kwon 1997)) that may function as synonyms or involve subtle or substantial differences in meaning (e.g., *amoral*, *immoral*, *non-moral*, and *unmoral*, cf. Zimmer (1964)). Some dictionaries provide separate entries for all such variants, others only for a selection of them. Much of the information on negated adjectives and the semantic distinction between variants with different prefixes – even in dictionaries that have some basis in corpus data – has been taken over from traditional sources to provide information on the perceived correct usage (Kwon 2001). This does not always coincide with the typical usage of such multimorphemic words in actual data. Which negative prefix attaches to an adjective depends on its general productivity, its individual general meaning, the adjective type, and etymological constraints. The specific prefix meaning also has to be seen in connection with the morphemes of the adjective stem and, if applicable, their suffixes.

Prefixed negated forms may have a particular meaning and not always simply mean the opposite of their unprefixed base (e.g., *unalive* is not primarily used to mean "not alive" or "dead," but "lacking in vitality" and "being not fully susceptible to something," cf. its OED entry). The prefix *dis-*, for instance, like other negative

 ISSN 2453-8035

prefixes whose roots go back to independent words that were not negation particles, often adds a specific meaning to English adjectives. It was borrowed under the influence of Latin and French (*des*-) during the Middle English period and is related to *bis* from Greek with the meaning "two," "divided in two." An adjective with *dis*- can simply mean the opposite of its base (*discontent*), but typically *dis*- also adds the meaning "away from," "apart" or "lacking a (valued) quality." It may also refer to a reverse state or process when attached to adjectives or to verbs from which adjectives are derived (Dixon 2014: 74, e.g., disconnected). It has a different origin than *dys*-, which is found in many medical terms (from Greek with the meaning of "ill," "bad," "abnormal," e.g., *dysfunctional*, *dysexecutive*). In Hulse's study (2010: 187) on the productivity of negative prefixes in the British National Corpus (BNC), hapaxes with *dis*- occurred only rarely in the language of the natural sciences and more frequently in areas that are more likely to influence non-specialised language use. The negative prefix *de*- means "to reverse an action" and can occur in adjectival participles derived from verbs typically ending in *-ize/-ise*, *-ate*, or *-ify* (*desaturated*, *dephosphorylated*). Certain types of synthetically negated adjectives may thus add additional nuances and a finer adjustment of information than negation with *not*. English negative prefixes can have a meaning of negation in connection with reversal of direction or action, inferiority, insufficiency, badness, wrongness, over-abundance, pejorative indication, opposition, or removal (Joshi 2020: 86).

The etymological origin of the stem of an adjective (and to a certain extent also the type of its suffix if it has one) plays an important role in the selection of the prefix. Some borrowed negative prefixes such as *a(n)*-, initially only attached to English words of Greek origin (e.g., *asymmetrical*), remain rare word formation elements in English (Funk 1971: 368), but have been shown to be much more important for the natural sciences than for other domains. They occur primarily in scientific and technical terminology, although some have spread to general language (Hulse 2010: 96). It has been suggested that rare negative prefixes such as *a(n)*- are more stylistically marked

 ISSN 2453-8035

and can be expected to be replaced by less marked ones over long time periods, as in a gradual replacement of *atypical* by *untypical* during the 20th century, a development which Bauer (2001: 50, 137) observed in general English language data. We cannot confirm this observation by Bauer with regard to the developments observed in scientific writing. *Untypical* was used before *atypical* in the RSC from the 1860s onwards, but *untypical* as a less formal variant remains less frequently used in scientific articles throughout LModE and PDE than *atypical*.

Negative affixes are usually attached to positive bases and not to evaluatively/emotively negative adjectives denoting some negative or undesirable property (cf. Horn 1989: 274-275; Zimmer 1964: 15). For those adjectives, negation with *not* is more typical. Negative prefixes cannot be easily stacked, but Funk (1971: 368) emphasises the ability of *non-* to attach to adjectives with a negative prefix, e.g., *in-* or *un-*. Observations in the BNC and the Corpus of Contemporary American English confirm the low productivity of negative prefixes in combination with evaluatively negative adjectives (exceptions may be found, e.g., *undangerous*, *undark*, *unbroken*) and of multiple negative affixes in adjectives (De Clercq & Vanden Wyngaerd 2019: 426). Lexemes that contain two negative prefixes can be occasionally found in the RSC, for example participial adjectives with either *un-* or *non-* and *dis-* or *in-* (e.g., *undisrupted*, *undisinfected*, *undischarged*, *non-discharging*, *non-independent*).

One advantage of using affixes for adjective negation instead of negation with *not* is the possibility to derive forms in other word classes with the same base and prefix (e.g., *non-uniform*, *non-uniformly*, *non-uniformity*) and to use such words throughout texts as a means of lexical cohesion. Another advantage is that affixal negation allows the construction of affirmative sentences involving semantically negative notions by avoiding more explicit clausal negation that might have different effects on the reader (Joshi 2012: 52-53). As affixal negation leads to sentences that are affirmative in

 ISSN 2453-8035

nature, one might argue that morphological negation is not *true* negation, and there is some discussion on which type of negation is the strongest or less ambiguous (Hulse 2010: 32). An important aspect of using affixes is economy of expression. Adjectives negated with prefixes contribute to achieving economy in terms of the number of words used (Joshi 2012: 52).

The following section provides an overview of ***un-*** and ***non-,*** the negative prefixes we focus on in our corpus analysis.

2.2 Un- and non- in negated adjectives

Un-$_1$ as a prefix of Germanic origin is the most productive negative prefix throughout the history of English with its basic meaning comprising negativity ("not," "opposite") and an additional pejorative meaning such as "lacking a valued quality" (Dixon 2014: 74; Kastovsky 2006: 237). From a historic perspective, it is of different origin and has a more general meaning than the less productive negative prefix *un-$_2$* expressing the reversal or cancellation of an action in verbs such as *unload*, *untie*, *unlock* and their derived adjectival participle forms. Bauer & Nation (2020) pointed out that *un-*, which can attach to adjectives and verbs with these different meanings, is potentially ambiguous, e.g., in *unclothed* (*un-$_1$*), which may mean "not clothed" or "having had the clothes removed" (*un-$_2$*), or in *unlocked*, which can mean "not locked" (*un-$_1$*) or function as a participle derived from *unlock* (*un-$_2$*). According to the OED entry for this prefix, the productivity of formations in *un-$_2$* has declined particularly since the 17th century as the prefix *de-* has become more frequent in verbs with reversive meaning. In the following discussion and analyses, we use *un-* to refer primarily to the most productive morpheme *un-$_1$*. *Un-* can be attached to both Germanic and Romance stems and to simple adjective roots (*unkind*, *uncalm*), but also to more complex adjectives derived from nouns and verbs (*unsuccessful*, *unavoidable*). It is also frequently affixed to participles functioning as adjectives (*unchanging*, *unpublished*). *Un-* can easily occur with a variety of descriptive, qualitative adjectives that may be gradable and that

 ISSN 2453-8035

can be modified by adverbs such as *very* and *rather* or used in comparative or superlative structures.

There is a certain historical competition of *un-* with *in-* that has received some attention in the research literature (between the 15th and 17th centuries largely a matter of choice according to the OED, cf. also Kwon 1997). According to Jespersen (1917: 139-140) *un-* is preferred before shorter, *easy* words and when a word has a native ending, and *in-* before longer ones and those of a more learned nature. *In-*, a borrowed prefix (*inanimate*, *inefficient*), and its phonologically conditioned allomorphs *ig-*, *il-*, *im-*, and *ir-* involving assimilation depending on the initial consonant of the adjective bases (*ignoble*, *illegitimate*, *impossible*, *irregular*) primarily attach to Latinate words (*impure* vs. *unclean*, *illegible* vs. *unreadable*). In most cases, the replacement of one negative prefix by another is not possible *(*insuccessful*, **unadequate)*. However, in Early and Late Modern English when many negated adjectives were newly coined, *un-* and *in-* variants were introduced and sometimes continued to coexist for a while or have acquired particular connotations (cf. OED entry on *un-*, prefix$_1$). In cases of a prefix variation in adjectives with identical meaning, there is typically a dominant variant in Modern English (e.g., *inadvisable* is preferred to *unadvisable* in our data; the last usage of the form *unadvisable* in the RSC occurs in the 1960s). *Un-* has spread to certain words that were initially formed exclusively with *in-*, especially in cases where the form with *in-* has acquired a more specific sense (e.g., *inhuman* vs. *unhuman*, cf. Hulse 2010: 27). *In-* is used in various multimorphemic borrowings from Latin and French. Although *independent*, *impossible*, and *irregular* are the most frequent examples of all synthetically negated adjectives in the RSC, this prefix is not very productive in PDE data compared to *un-* and other negative prefixes (cf. also Hulse 2010: 8).

Non- as a borrowed affix has become one of the most productive negative affixes for adjectives in PDE. Initially, it was mainly used in specialised legal, scientific, or technical contexts. It has been added to adjectives with increasing productivity and

fewer restrictions from the 17th century onwards and has started to occur frequently in specialised and non-specialised registers. It also creates opposites for nouns (e.g., *non-conductors*). It occurs in adjectival participles (*non-thermalized*, rsta_1966_0057), and it is rarely attached to verbs as in *non-dimensionalize* (RSC 6.0.1 text ID: rspa_1969_0040). There is no agreement in the literature on the word class of forms with *non-* in examples such as *non-stop flight* or *non-slip soles.* Their first constituent may be analysed as a prefixed adjective or a compound adjective where *non-* is attached either to a verb and changes its word class, to an adjective created by conversion, or to a shortened form, e.g., created by a backformation from an adjectival participle as in *non-stopping flight*. Therefore, there are also different opinions on the question of whether the entire constructions function as nominal compounds or phrases (Allen 1978; Štekauer 2000: 211).

Like *un-*, *non-* can be used to create semantically transparent negated adjectives. Dictionaries do not list all of them as entries. *Non-* has the meaning of "not a member of a specified class" and can have scope over single words and complex constituents (e.g., *non-[profit-making] company*) of various etymological origins (Dixon 2014: 74, 92). *Non-* differs from *un-* in that it predominantly expresses a binary contrast (e.g., *non-scientific* vs. *unscientific*, cf. Quirk et al. 1985: 1540). Semantically, *non-* is the prefix that is most similar to the black and white nature of negation with *not.* Adjectives with *non-* are typically classifying adjectives that are contradictory antonyms of their corresponding non-negated forms and express contrast without gradability (Funk 1971: 372, e.g., *green and non-green algae*, rstb_1986_0042). Exceptions of gradable adjectives with *non-*, e.g., modified by *very*, are rare (e.g., *a very nonlinear world*, rstb_1994_0162, *a very non-trivial problem*, rsta_1989_0083). *Non-* is generally less emphatic than *un-* and primarily confined to descriptive, i.e. *neutral*, terms (Zimmer 1964: 33-34). It generally adds a less pejorative meaning than *un-*.

 ISSN 2453-8035

Non- has a hyphenated and a non-hyphenated variant (*nonconducting power/non-conducting power*). Due to the optional hyphen as a visible boundary between the morphemes, *non-* has more flexibility than *un-* with regard to its possible scope over single words or complex constituents. *Un-* is rarely used as a hyphenated variant (e.g., *ice-like and un-ice-like species*, rspa_1958_0206). *Non-* can attach freely to native bases (*non-flesh-forming material*, *non-singing small species*) and is flexible as a marker of various words and word-like units from different word classes. The patterns it attaches to in scientific language are quite diverse (*non-insulin dependent diabetes*, *non-vacuum jacketed calorimeter*, *non-single-unitary symmetric functions*, *non-species-specific inhibition*, *non-electro-negative bodies*, *non-heat-evolving medium*, *non-uniformly heated portion*). It can also be attached to premodifying items with capital letters, such as eponyms or acronyms (*non-Euclidian geometry*, *non-NGF dependent neurons*, *non-GI visceral organ*), and to other forms that would be difficult to be negated by other prefixes, e.g., colours and numbers (*non-yellow homozygotes*, *non-zero current levels*). *Non-* is potentially less ambiguous than *a(n)-*, *de-*, *dis-*, *in-* (*il-*, *im-*, *ir-*) and *un-* as it rarely occurs in non-negated adjectives that start with the same sequence of letters (e.g., *nonic*).

The next section will give an overview of phrasal and clausal constructions that contain adjectives and the negation marker *not*.

2.3 Analytic negation with 'not' + adjective

This section discusses some aspects of non-affixal negation by the addition of the negative particle *not* (cf. Biber et al. 1999: 158). It is debatable whether the world can actually be described without the use of the word *not* as Russell suggested (1948: 520, cited in Horn 2001: 50). The scope of negation with *not* may extend from the negative marker to include the entire clause. Negation with *not* produces a contradictory opposition that is binary and non-gradable. It may co-occur with main or subclause patterns that involve adjectives which are not synthetically negated. *Not* also has the

 ISSN 2453-8035

contracted form *n't* that can be attached to auxiliary verbs, but this form does not occur in the scientific articles in the RSC. Using the negation marker separately makes emphasis on *not* possible. The negation marker may, for instance, occur in the topic position to receive a special focus (*Not identifiable with any prominent alias is the largest peak in the PA spectrum.* rsta_1976_0020). Adjectives often carry a new information load. Using an adjective in a clause-final position and not as a noun premodifier puts it into a focus position. A clause-final occurrence of a predicatively used adjective combined with the function word *not* may lead to a slightly lesser informational peak than a morphologically more complex, synthetically negated adjective in this position.

The scope of negation with *not* may also be restricted to individual adjectives or adjective phrases without making the clause negative, but this type of local negation is beyond the scope of our analysis (e.g., *a not very successful approach*). *Not* can occur with synthetically negated adjectives in double negation (litotes) in registers that are rich in figures of speech. Litotes can also be found in scientific contexts (*the walls are not non-reflecting*) or function as a politeness marker in academic discourse (*it is not disrespectful to say …,* rsbm_1944_0016) where *not* may reduce the negative force of a negative adjective. In this article, we are specifically interested in structures with *not* + positive adjectives.

The question of whether synthetically negated adjectives and constructions with *not* and corresponding non-negated adjectives are equivalent in meaning cannot be answered straightforwardly. Due to the nuances in meaning that a negative prefix may add to an adjective, there might be subtle differences in meaning between prefixed negated adjectives and constructions with the negation marker *not*. Some adjectives cannot be negated synthetically, and a few adjectives with negative prefixes do not have an independent non-negated base in PDE or their base is not easily recognizable (e.g., *inept*, *inert*, *uncouth*, *unkempt*, cf. Bauer & Nation 2020; Horn 1989: 275).

 ISSN 2453-8035

According to Calude & Bauer (2021), there is a sequence of degrees for scalar and stance adjectives, e.g., *likely – not unlikely* (= rather / fairly likely) – *not likely – unlikely*. Therefore, *likely* and *not unlikely* as well as *not likely* and *unlikely* are not perceived as synonymous by them. Also, *not happy* is not necessarily *unhappy* and the adjective *invalid*, for instance, has different senses of which *not valid* is only one possible meaning. Nevertheless, synthetically negated forms can be identical in meaning or at least approximately synonymous with adjectives negated by *not*. There seems little semantic distinction between the types of negation in examples such as in (1) or between the variants of (2).

(1) a) *A bachelor is a man who is not married.*
    b) *A bachelor is man who is unmarried.*
    c) *A bachelor is an unmarried man.*
(2) a) *The remark was inappropriate.*
    b) *The remark was not appropriate.*

In fact, very few theoretical, corpus-linguistic or psycholinguistic studies compare the functions of different types of negated adjective constructions, such as affixal vs. non-affixal patterns. Tottie (1980; 1999) is one of the few who discusses both patterns. She contrasts adjective pairs from Middle English and PDE texts such as *unprofitable* vs. *not profitable* with regard to the issue of semantic equivalence or difference of such pairs and the productivity of selected affixes. Tottie points out some stylistic and syntactic reasons for opting for one of these variants. Her hypothesis was that the variants represent different systems in almost complementary distribution, which is why she assumes that they remain stable over time (1999: 247, 258-262). Her results partly confirm this and show that certain constraints on affixal and non-affixal negated adjective constructions remain relatively similar from Middle English to PDE. However, she also notes that sampling restrictions led to sparsity of data for specific constructions and the Late Middle English and PDE data have different rhetorical

styles.

Aina et al. (2018) used web-crawled corpora and built a distributional semantic model to investigate the relation between adjectives negated with *not* (e.g., *not cold*, *not happy*), the adjectives themselves, their lexical antonyms (e.g., *hot*), and their morphologically related antonyms derived by adding a prefix (e.g., *unhappy*). Their findings on the respective contexts of use suggest that adjectives negated with *not* are typically more similar to the adjectives themselves than to their antonyms, but this effect seems to be weaker when antonyms are derived via negative prefixes.

2.4 Previous research on the mental processing of negation

Farshchi et al. (2020) state that research focussing on the processing of negation has primarily targeted analytic forms of negation, and the processing of prefix negation has been given little attention in empirical analyses. The results from psycholinguistic studies on prefix negation, which tend to be based on short sample sentences considered in isolation, are not in agreement with each other. As there is little research on how prefixed patterns are processed or how they are comprehended in comparison to other forms of negation, it is not clear yet whether the processing of synthetically negated adjectives is similar to that of adjectives negated with *not* or non-negated adjectives. Nevertheless, Farshchi et al. (ibid.) were among the first to conduct a psycholinguistic study on the processing of synthetically negated adjectives, the negation of adjectives with *not* and non-negated affirmative adjectives. Their experiment was based on a set of generated, relatively long sentences that resemble actual language use in news media texts. The results suggest that generally both types of negation involve a higher processing cost for working memory than non-negated adjectives. Farshchi et al. (ibid.) only partly found evidence in support of existing models on the processing of negation, and they also note that not all instances of negation function simply as rejections of information and may also convey attitudinal aspects.

 ISSN 2453-8035

2.5 Previous corpus studies on synthetic and analytic negation

This section presents an overview of several previous corpus studies on selected aspects of negation in English. There are corpus studies on various negation patterns, but only a few on synthetically negated adjectives. The corpus-linguistic studies that investigate negative prefixes typically address the distribution of selected prefixes and the question of how to measure their productivity, as studies from other linguistic fields also did, for instance, via psycholinguistic experiments (e.g., Baldi et al. 1985). Various often-cited works on negation in general and on adjective negation in particular were published as early as several decades ago (e.g., Funk 1971; Jackendoff 1969; Zimmer 1964) and are, for obvious reasons, not yet based on actual usage in larger corpora. Jiménez-Zafra et al. (2020) list a few corpora annotated with various means of negation, mostly syntactic negation markers. A difficulty they noticed was that potential negation markers can be ambiguous and do not always function as negators. Tottie (1991) compared negation with *not* and *no* in corpus data. She also analysed synthetically negated adjectives in Middle English data from the Helsinki Corpus and compared them to a PDE sample (several hundred thousand words in total, 1999). The samples included text types such as romance, drama and private letters. She focused on *un-, in-* and *dis-* and did not study the usage of *non-* as a negative prefix as it was not freely used yet in Middle English (only in a few technical and legal terms from the end of the 14th century onwards) and it only became productive from the mid-17th century onwards.

In Kjellmer's (2005) analysis of affixal negation of adjectives in the PDE Cobuild Direct Corpus including both speech and writing, *un-* occurred in 43% of the negated adjectives, *non-* in 22%, *in-* in 12%, and *dis-* in 6% of the negative prefixes that were studied, while 17% of the negated adjectives in the data had prefix variation. A high number of synthetically negated adjectives had no attested corresponding positive forms occurring in this dataset (e.g., *disadvantaged*, **advantaged*). Prefix variation most typically occurred in adjectives of Romance origin, sometimes with semantic

 ISSN 2453-8035

differences between them, and in most cases included a form with *non-* (e.g., *inorganic – non-organic – unorganic*; *disfunctional – non-functional*). He also found some variation among Germanic adjectives (e.g., *nondrinkable* vs. *undrinkable*). A brief look into our RSC data also shows that there are several variants as well (*unchemical – non-chemical*; *unmagnetic – non-magnetic*), possibly many *non*-words have early *un-* variants that did not become the dominant form. However, in scientific texts – particularly from a diachronic perspective – differences in meaning or synonymy might be more difficult to detect than in general PDE data.

A detailed synchronic analysis of the productivity of five negative prefixes occurring in adjectives and other word classes has been performed by Hulse (2010). She created a database of all negatively prefixed open-class words based on the BNC and compared prefix productivity and domain specific preferences. Different variables such as types, tokens, and hapaxes were examined to compare their role in various formulae used to measure morphological productivity. She found that in the PDE dataset *non-* is the most productive of the examined prefixes, while *in-* is the least productive. The application of productivity formulae involving the words that occur only once in a corpus (cf. also Baayen & Lieber 1991) led to the observation that the prefixes that turned out to be generally the most productive in written English, were not very common in spoken language, while the least productive ones in writing were among the most common ones in speech. An explanation for this might be that, in spoken language, negative prefixes occur primarily in familiar, high frequency words, but not in many infrequent words and new coinages (Hulse 2010: 187). In our case of diachronic, highly specialised data, we are aware of a still notable frequency of optical character recognition (OCR) errors among hapaxes despite the application of post-correction techniques. It is more difficult to automatically process large amounts of historical electronic texts across a time span of more than 300 years with high accuracy compared to data from only contemporary texts. It is also more difficult to linguistically annotate specialised scientific corpora than general language corpora. It would

 ISSN 2453-8035

therefore be only partly possible to replicate previous studies such as Kjellmer's (2005) and Hulse's (2010) by applying their methods and queries to our datasets. When working with variables such as **types** (i.e. the distinct word forms or lemmas found in a corpus) or **hapaxes** (items occurring only once in the corpus), differences in corpus size, annotations layers, query options, and corpus-specific error types, e.g., in POS tagging, need to be taken into consideration.

In their attempt at building a dictionary of affixal negations, Van Son et al. (2016) noted that if one wanted to consider all types of affixal negation, this would be rather difficult to detect automatically without a substantial false positive rate. Blanco and Moldovan (2011) stated that no simple search could unequivocally distinguish between negated words such as *ineffective* and other words that happen to begin with the same letters. The problem might be partially solved by checking if the word is still valid when the prefix is removed, but this method would falsely classify *informed* as negation because *formed* is still a valid word.

Biber et al. (1999: 159) used the Longman Spoken and Written English Corpus and found that negative words (*not* [so called *not*-negation], *no*, *nothing*, *nobody*, *nowhere*, etc. [so called *no*-negation]) in general are more common in speech than in written language as conversation is rich in verbs, and clausal negation is often tied to the verb. PDE academic prose has been identified as a register with a lower number of negative words than conversation, fictional prose, and news reportage. However, *not*-negation is most frequent in the two rather different registers of conversation and academic prose (ibid.). An explanation for this could be that these two have a higher proportion of negative clauses than other registers and they involve argumentation, and consequently greement and disagreement. Additionally, various verbs collocating strongly with *not* are frequent in both, e.g., mental verbs such as *know* or *think*.

 ISSN 2453-8035

There are several corpus studies that look cross-linguistically at selected negative prefixes in comparable corpora or parallel corpora. It has been confirmed that the productivity of negative prefixes varies in a comparable way across registers in English and French (Lefer 2012: 7), where prefixation is more productive in news editorials and scientific articles than in fictional texts. Another finding in translation corpora was that translators often resort to non-morphological translations when translating from English into a Romance or other Germanic language, even if a prefixed equivalent with identical meaning is attested and frequent in the target language (Cartoni & Lefer 2011: 813; Lefer 2012). In their work on learner corpora with texts from learners of English with different native languages, Gilquin & Lefer (2017) noted a general tendency among learners to underuse morphologically derived adjectives and to overuse syntactically negated adjectives. Translators and learners of English as well as native English speakers in specific registers seem to prefer morphologically less complex forms and a lower degree of phrasal complexity features.

On the basis of the challenges that were identified related to precision and recall in previous studies when applying queries for a great diversity of negation phenomena and on the basis of Kjellmer's (2005) finding that *un-* and *non-* are among the most productive negative prefixes, we decided to restrict our analysis of synthetically negated forms to adjectives with two negative prefixes, *un-* and *non-,* as relatively frequent negation markers and rather unambiguous sequences of letters – and to contrast it with *not*-negation of adjectives.

3. Hypotheses on linguistic effects of affixal vs. non-affixal adjective negation in scientific language

We focus on the diachronic development of affixal vs. non-affixal adjective negation. In particular, we consider the prefixes *un-* and *non-* paired with adjectives in comparison with the *not* + adjective forms for the reasons explained above. Our theoretical framework is based on the register theory (Biber 1995; Halliday 1985;

 ISSN 2453-8035

Quirk et al. 1985) as well as the information theory (Shannon 1948, cf. Section 4.1) adopting a communicative perspective. Our general question is whether academic writing overall has become more informationally dense over time. Given shared knowledge among academics, more condensed synthetic forms contribute to higher economy of expression (i.e. use of shorter, less explicit linguistic encodings), functioning as register-specific strategies used by writers to manipulate the information density distribution of particular linguistic structures with regard to the expectations and mental processing capacity of their addressees.

We assume that scientific argumentation needs a certain amount of disagreement markers including *not*. In scientific writing, we expect to find a high number of negated constructions with non-scalar adjectives expressing contrasts and non-evaluative meaning transfer in a descriptive, informative, and non-pejorative way. Negated adjectives also play an important role in conventionalizing collocations with nouns and in term formation (cf. Degaetano-Ortlieb & Teich 2019). Especially when the prefix *non-* is used and when *un-* or *non-* are attached to adjectival participles, clausal variants with *not* + the corresponding positive adjective or adjectival participle will be potentially equivalent in meaning (e.g., *uncontaminated water* [rspa_1925_0127] = *water that is not contaminated).* Some scholars have suggested that affixal negation is generally more likely to be found in written, formal language, whilst forms of negation with *not* are more typical of spoken language (cf. Tottie 1980: 104 or the study on the BBC by Hulse 2010: 188). We therefore expect to see a decreasing use of negation with *not* and an increasing usage of affixal negation in scientific writing over time. Particularly scientific papers before the Present-Day English period will probably be characterised by a more involved style and a higher proportion of analytic negation markers. Thus, we pose the following hypotheses:

 ISSN 2453-8035

H1: *Increase in the number of prefixed negated adjectives*: In scientific English, prefixed adjectives with *un-* and *non-* will increase in frequency over time compared to the *not* + adjective form.
H2: *Synthetic forms are more informationally dense*: Adjectives prefixed with *un-* and *non-* as shorter, more compact forms of encoding will show higher information density than adjectives negated with *not*.

From a cognitive perspective, one of the reasons to favour more informationally dense forms, as in Example (3a) from the RSC, might be to avoid additional memory workload caused by complex syntactic patterns (i.e. long dependency length with regard to the distance between a linguistic head and its dependents, see Gibson 2000; Juzek et al. 2020). The generated alternative (3b) results in a longer variant with a relative clause and an analytically negated adjective that is further away from the head noun than in (3a).

(3) a) *Those experiments apply only to the heat evolved from a non-luminous source.*
b) *Those experiments apply only to the heat evolved from a source which is not luminous.*

Regarding the evolution of modern science and the development of scientific language, scientific English reflects processes of specialization resulting in the ongoing creation of new lexemes (Degaetano-Ortlieb & Teich 2019), many of them formed according to regular word formation rules and processes. This process of specialization has also contributed to an increasing need for adjectives that denote properties, including logically complex properties such as "negative properties" (Zangwill 2011). Additionally, conventions have formed such as the use of formulaic expressions or terminology, and stylistic preferences have contributed to a shift towards fewer oral and more literate features (Biber & Finegan 1997; Degaetano-Ortlieb et al. 2019; Degaetano-Ortlieb & Teich 2016). Such specialization and conventionalization

 ISSN 2453-8035

processes seem to act as a balancing mechanism allowing an optimal code of scientific communication (Degaetano-Ortlieb & Teich 2019). Thus, we will also consider how affixal vs. non-affixal adjective negation might be affected by specialization and conventionalization processes.

4. Methodology and corpus data

4.1 Tracing the development of informationally dense structures

Besides comparing frequency distributions over time, we use surprisal as an information-theoretical complexity metric to measure the amount of information particular patterns carry and a predictor of cognitive load (Hale 2001; Levy 2008). Surprisal is the context-specific predictability of linguistic items, e.g., how probable the occurrence of a particular word is after a specific preceding context. For example, given the three words *Jane bought a*, surprisal indicates how predictable the following word *book* would be in comparison to how predictable *book* would be after *Jane read a*. As *book* more frequently occurs with the verb *read* rather than *bought*, the word *book* would be more predictable given the context with the verb *read*. Formally, surprisal *S* is calculated by the $-\log_2$ probability *p* of a *word* in *context* denoted by the formula $S(\text{word}) = -\log_2 p(word|context)$.

In accordance with previous work on the RSC (Degaetano-Ortlieb & Teich 2019), we consider a lexical trigram context window of three tokens (including words and punctuation marks) to the left of each word. If a word has high surprisal, its predictability given the previous context is low, while words with low surprisal are more easily predictable given their previous context (consider again *Jane bought a book* vs. *Jane read a book*). High surprisal and low predictability indicate a large amount of information, while low surprisal and high predictability indicate a small amount of information. We will use surprisal to test whether prefixed adjectives with *un-* and *non-* are more informationally dense than the *not* + adjective form and to test whether the amount of information in these patterns has changed over time. For the

purpose of comparison of surprisal over time, surprisal models are sensitive to time periods. For example, to compare periods of 50 years, surprisal models are based on probabilities obtained from 50-year periods in the corpus data. This is an important step to ensure comparability when analysing change in language use, as the probabilities of word occurrence will change over time based on the contexts words appear in.

Surprisal has been shown to be proportional to cognitive effort, i.e. words with high surprisal indicate higher processing effort for these words, while words with low surprisal are easier to be processed (cf. Hale 2001; Levy 2008). This becomes particularly clear when considering words in conventionalised lexico-grammatical contexts vs. specialised terminological contexts (cf. Degaetano-Ortlieb & Teich 2019): in a formulaic pattern such as *it is not possible to*, the adjective *possible* is quite predictable in comparison to the adjective *non-luminous* in the sequence *evolved from a non-luminous source.* Thus, the comparison of surprisal of adjectives across local linguistic contexts (such as three preceding tokens) and over time will allow us to (1) examine whether the prefixed adjectives with *un-* and *non-* (we will refer to them also as *un-* and *non-*adjectives) become more informationally dense over time in comparison to the *not* + adjective form and (2) whether those forms are subject to diachronic conventionalization and specialization processes.

4.2 The Royal Society Corpus

The data set used for our analyses is the Royal Society Corpus (RSC 6.0.1) available from the Saarbrücken CQPweb interface. It is a well-curated corpus of scientific English covering approximately 350 years. Version 6.0.1 contains 47 837 texts and 295 895 749 tokens (of which almost 80 million tokens from more than 250 years are part of the freely accessible open corpus version). Table 1 presents an overview of the seven 50-year subcorpora in the RSC and their respective size.

 ISSN 2453-8035

Table 1. Size of the Royal Society Corpus 6.0.1

Time Period	Tokens
1650 (1665–1699)	2 582 856
1700 (1700–1749)	3 414 795
1750 (1750–1799)	6 342 489
1800 (1800–1849)	9 112 274
1850 (1850–1899)	36 993 412
1900 (1900–1949)	65 431 384
1950 (1950–1996)	172 018 539

The RSC texts are digitised versions of professionally published scientific journal articles, mainly from the "Philosophical transactions" and the "Proceedings" of the Royal Society of London from 1665 to 1996. Both journals were rather general in the early years and became more specialised in the mathematical, physical and biological sciences over time after they split into A and B series in 1887 and 1905 respectively. Word formation aspects in the language of these academic journal articles (e.g., with regard to combining forms and eponyms) have, for instance, been addressed in Menzel (2021) and Menzel & Degaetano-Ortlieb (2017). The following extracts in Fig. 1 and 2 illustrate the language used in such articles and contain various examples of negated adjective constructions, several of which are adjectival participles with a negative prefix, particularly in Fig. 2.

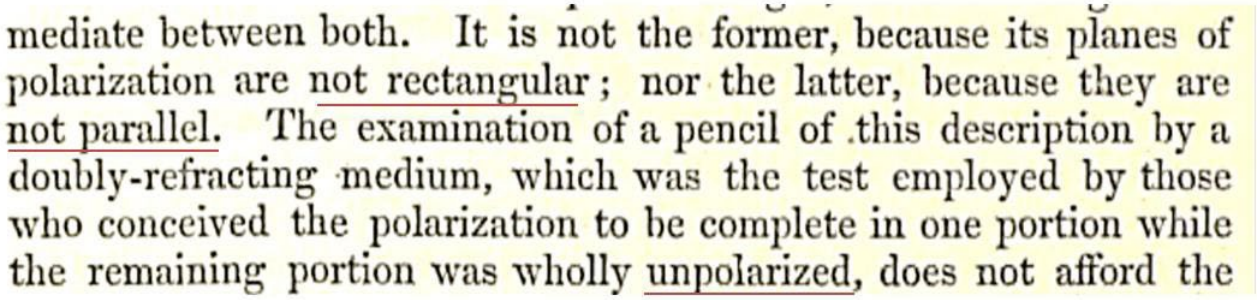
mediate between both. It is not the former, because its planes of polarization are not rectangular; nor the latter, because they are not parallel. The examination of a pencil of this description by a doubly-refracting medium, which was the test employed by those who conceived the polarization to be complete in one portion while the remaining portion was wholly unpolarized, does not afford the

Figure 1. Extract from "On the law of the partial polarization of light by reflexion ", D. Brewster, 1833, RSC ID: 107890

ISSN 2453-8035

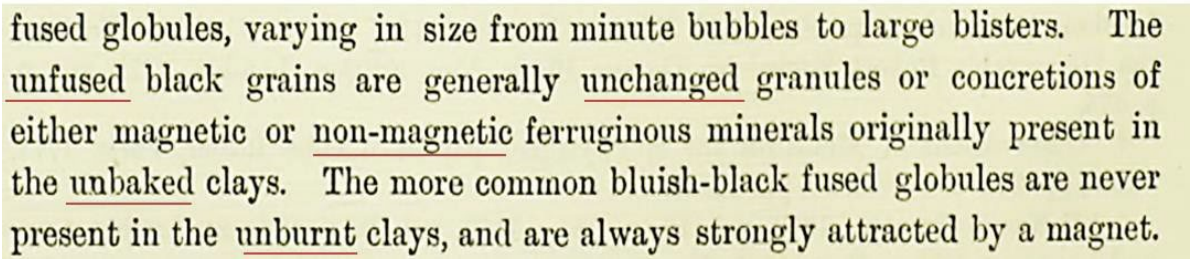
fused globules, varying in size from minute bubbles to large blisters. The unfused black grains are generally unchanged granules or concretions of either magnetic or non-magnetic ferruginous minerals originally present in the unbaked clays. The more common bluish-black fused globules are never present in the unburnt clays, and are always strongly attracted by a magnet.

Figure 2. Extract from "The magnetic materials in claywares", A. Hopwood, 1913, RSC ID: rspa_1913_0058

4.3 Selection and extraction procedure

In our diachronic corpus analyses the focus is on synthetically negated adjectives with *un-* and *non-* (we make no particular distinction between *un-*$_1$ and *un-*$_2$ in our queries as very few adjectives in the data fall under the latter category, cf. Section 2.2). These two prefixes are also much less ambiguous sequences of letters at the beginning of adjectives in contrast to other negative prefixes representing quite ambiguous sequences of letters that occur in many non-negated adjectives, for instance, *distributed, illustrious,* and *impending* do not contain the negative prefixes *dis-*, *il-*, and *im-* respectively.

For the sake of simplification, we also disregard the possibility here that the different patterns that we contrast with each other might sometimes represent variants with slight semantic nuances or may not be easily interchangeable in certain functions and contexts. They might also represent negation subtypes in connection with specific adjective constructions which are beyond the scope of our analysis and will not be searched for in our queries, e.g., negation with *not* when adjectives are premodified by adverbs in the RSC as in *"considerable portions of a ship's iron which are not permanently magnetic"* and contexts of double negation with *not* preceding adjectives with a negative prefix as in *"These cross-loads are not unfrequent in the mines on North Downs".*

We use **CQPweb**, an online corpus analysis system (Hardie 2012), to search the RSC for the three POS patterns: (1) adjectives prefixed with *un-*, (2) those with *non-*, and

 ISSN 2453-8035

(3) analytically negated adjectives with *not*. POS information in our dataset has been annotated using the Penn Treebank tagset. An alphabetical list of the POS tags from the tagset can be found online (Penn Treebank tagset, *s.a.*), and a more detailed explanation of the tags is given in Santorini (1990). We consider only words tagged as adjectives (JJ.*) in our queries. To increase the precision of the results, we restrict the query for analytically negated adjectives by excluding all words that contain special nonalphabetic characters arising from OCR errors. In the case of adjectives negated with *un-*, we exclude words with the prefixes *uni-* and *under-* from the query to enhance its precision considerably (although this also excludes occasional cases that might have been relevant negated forms). The queries for the adjectives negated with *non-* were refined by excluding special characters and possible Latin lexemes (e.g., *nonnihil*).

4.4 Analytical approach

First, we adopt a macro-analytical perspective to investigate general changes of frequency distributions as well as surprisal over time for the two negative prefix + adjective patterns and the *not* + adjective pattern. Second, we take a micro-analytical view to inspect the local linguistic context of all three forms and the possible impact the context might have on surprisal. For this, we inspect part-of-speech (POS) sequences preceding and following the three forms. Due to a high number of different POS tags in the original tagset, e.g., distinguishing between tense- and aspect-related information, in our analysis we disregard some of the more detailed POS distinctions of the original tagset that have little relevance to our purposes here. For instance, we summarise VBZ (verb, *be*, present) and VBG (verb, *be*, progressive) to VB, etc. In our case study, we also subsume all noun forms under N, all adjective forms under J and all adverbs under RB. Also, we summarise round brackets and quotation marks to PUNCT.

 ISSN 2453-8035

5. Analysis

5.1 Distribution of prefixed adjectives and not + adjective forms

Looking at the development of prefixed adjectives with *un-* and *non-* as well as *not* + adjective in the RSC, in line with our first hypothesis, we find that both affixal negations become more frequent over time, while negative non-affixal patterns with *not* slightly decrease (Fig. 3).

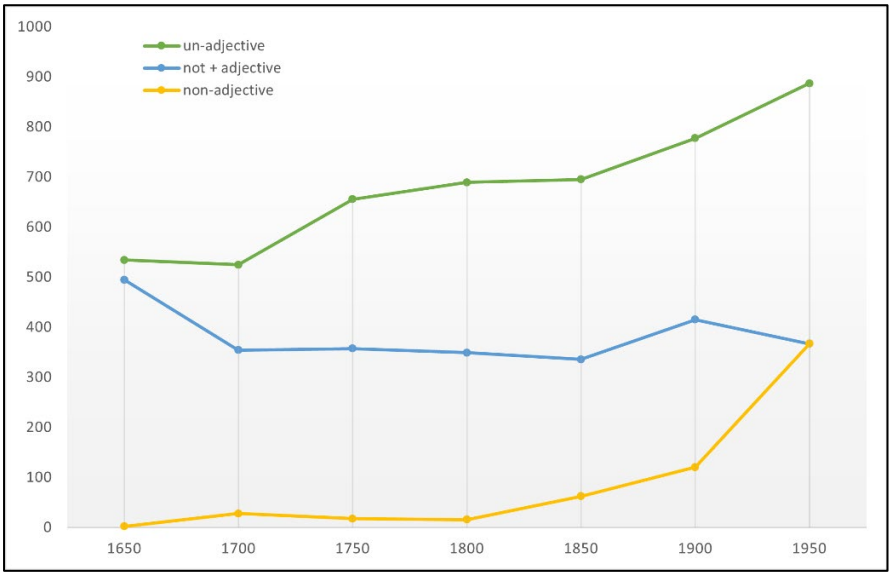

Figure 3. Relative frequencies (per million tokens) of negated adjectives with *un-* and *non-* and of adjectives preceded by *not* across 50-year periods in RSC 6.0.1 (i.e. 1700 covering 1700–1749, 1950 covering 1950–1996, etc., cf. Table 1)

Adjectives with the prefix *un-* almost double in frequency (~500 to ~900 instances per million tokens) and show a consistent increase in number over the whole observed time span. Adjectives with *non-* remain vanishingly infrequent during the 17th and 18th centuries to increase steeply between 1800 and 1996, especially in the last 50-year period. Adjectives with *non-* have become a popular word formation pattern in the more recent texts, while *un-* started out as a common negative prefix and has continuously become more frequent. *Not* + adjective, instead, shows an overall moderate decrease towards the end of the 19th century and a slight increase afterwards. At the end of the 20th century, adjectives with *non-* and the *not* + adjective pattern are equally frequent. We can confirm our first hypothesis that, over time, affixal negation

ISSN 2453-8035

of adjectives with regard to the selected two prefixes is used more strongly in English scientific journal texts, whereas analytic adjective negation has become less frequent.

5.2 Development of informationally dense forms

For our second hypothesis, we use surprisal to see whether synthetically negated adjectives are more informationally dense than analytically negated adjectives (Section 5.2.1), and whether the amount of information these forms transmit may change over time pointing toward processes of conventionalization or specialization. We analyse both the preceding and following contexts (Section 5.2.2).

5.2.1 Macro-analytical surprisal tendencies

In general, adjectives preceded by *not* have the lowest surprisal of all three negation patterns confirming our second hypothesis (compare Fig. 4 to Fig. 6). In addition, diachronically the surprisal of adjectives negated by *not* decreases despite their decreasing frequency over time (cf. Fig. 3). This indicates that adjectives negated by *not* start to require less processing effort over time, since they become more predictable. This might be counterintuitive given the fact that adjectives negated by *not* decrease with regard to their overall frequency. A plausible explanation for this development is that, over time, the contexts in which this negation pattern occurs become increasingly conventionalised, i.e. they occur with increasingly similar preceding contexts compared to earlier time periods (e.g., *"similar if not identical"*, *"n.s. = not significant"*), while other contexts are highly unpredictive with regard to this negation pattern (*"compound was not diazotizable," "Postosuchus are not crocodilian"*). So, the frequency with which the preceding trigram occurs with an exact word, in our case with a specific adjective, in relation to how often it occurs with other words determines the predictability of that word. For example, *"similar if not"* is always followed by *"identical"*, so the adjective is highly predictable. In contrast, *"compound was not"* occurs before many different adjectives, e.g., *hydrolysed*, *soluble*, etc., so this sequence is not very predictive of the adjective *diazotizable*, and hence

 ISSN 2453-8035

diazotizable has a high surprisal value in this context.

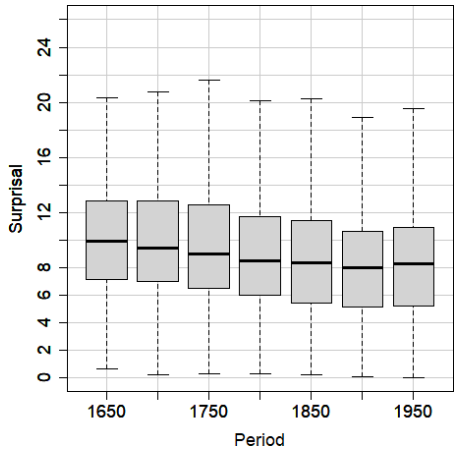

Figure 4. Surprisal of adjectives preceded by *not* across 50-year time periods in RSC 6.0.1

Adjectives prefixed with *un-* (Fig. 5) show median surprisal values that are higher than the surprisal values of adjectives after *not,* but lower than those of adjectives with the prefix *non-* (Fig. 6). Over time, the median surprisal for *un*-adjectives decreases very slightly. The boxplots show that 25% of the surprisal values occupy a lower and decreasing range of values in the last half of the 20th century compared to the 17th century, while 75% of the surprisal values between 1950 and 1996 occupy a larger range than those in 1665–1699. This shows that in 1950–1996, the surprisal values are more concentrated in the lower area between 9 and 13. However, the increasingly large whiskers in the bar plots of the period of 1950–1996 show that in this period surprisal could stretch from very low to increasingly high values pointing to *un*-adjectives occurring in very general, nondeterministic contexts (*"that of the unpierced"*, *"the nature of undead"*) as well as more conventionalised and more frequently occurring contexts (e.g., *"in scrapie-infected and uninfected")*. Note that some contexts occur among the frequent ones if they are part of longer constructions that are quoted verbatim throughout different publications such as references to article titles. The use of quotations and references has become particularly frequent in the last 50 years of the data.

 ISSN 2453-8035

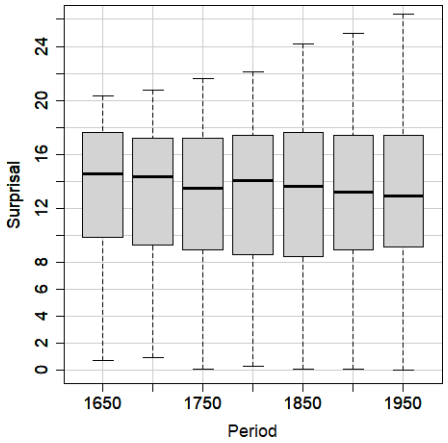

Figure 5. Surprisal of adjectives with *un-* across 50-year time periods in RSC 6.0.1

For *non*-adjectives (Fig. 6) surprisal is the highest of all the three negation patterns.[2] Interestingly, despite the much lower frequency of *non*-adjectives between 1650 and 1850, surprisal values start out with lower medians in the first two periods and increase towards 1750. Afterwards, the surprisal medians stabilise approximately between 16 and 18. The boxplots indicate that in the first 100 years when adjectives with *non-* are generally quite infrequent, they are used in rather similar contexts (indicated by lower surprisal), while in the later periods their frequency soars and with it the variety of contexts as well increments (indicated by higher surprisal). In addition, in the latest period (1950–1996) the surprisal median drops again, possibly indicating that some of the uses of the relatively new form of negation settle in conventionalised contexts. Also, the surprisal range is relatively wide for adjectives with *non-* indicating both the use of highly unpredictive contexts (*"Self-order in flexible non-mesogenic"*) and increasing processing effort on the adjective, and more predictive contexts (e.g., *"useful ores and non-metallic," "the mimetic and non-mimetic"*) where less processing effort is needed.

In summary, the surprisal analysis has shown that, indeed, the analytic pattern *not* + adjective becomes easier to process due to increased predictability of the adjectives within their preceding contexts. We assume that this higher predictability can be attributed to the pattern occurring in increasingly similar contexts. The constant use of the function word *not* also has a certain influence on the overall surprisal of the trigrams

 ISSN 2453-8035

preceding the adjectives in this negation pattern. The two synthetically negated adjective types develop in divergent directions. *Un*-adjectives tend to require slightly lower processing effort over time, which can be attributed to the increasing frequency alongside rather stable contexts of *un*-adjectives, while *non*-adjectives overall become harder to process, possibly due to a diversified use in varied contexts.

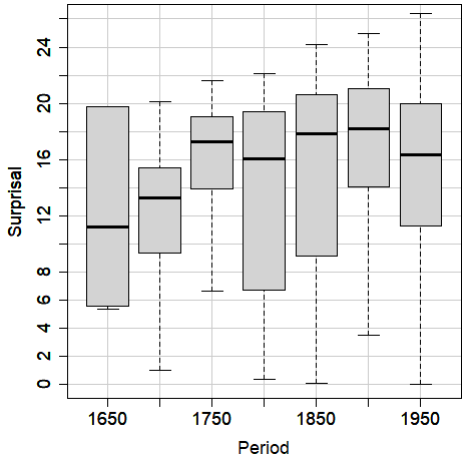

Figure 6. Surprisal of adjectives with *non*- across 50-year time periods in RSC 6.0.1

5.2.2 Micro-analytical inspection of preceding and following contexts

To find out why exactly surprisal values of the adjectives in the synthetic and analytic negation patterns diverge, we conduct a micro-analysis of the preceding contexts of the negated adjectives that we identified with our queries. To obtain an overview of their grammatical syntagmatic contexts, we extract the preceding **POS trigrams** concentrating on the most frequent ones in each time period. We further extract the most frequent preceding **lexical trigrams**, which ultimately form the basis of the surprisal score calculated for each adjective (cf. Section 4.1).

5.2.2.1 Not + adjective

We first look at the most frequent POS trigrams preceding the pattern *not* + adjective. For this, we extract all trigrams that are found among the five most frequent POS trigrams in at least one of the 50-year periods of the corpus and highlight the five most frequent ones of each period with a black border (Fig. 7). We see that three trigrams

 ISSN 2453-8035

become increasingly prominent over time: [N-N-VB: noun – noun – verb *be*] (green), [J-N-VB: adjective – noun – verb *be*] (dark blue) and [SENT-PP-VB: full stop – personal pronoun – verb *be*] (light blue). The first two represent parts of complex noun phrases followed by the copular verb *be*. The third pattern represents sentence-initial evaluative patterns (e.g., *". It is not possible"*, *". It is not correct"*) of a highly conventionalised form.

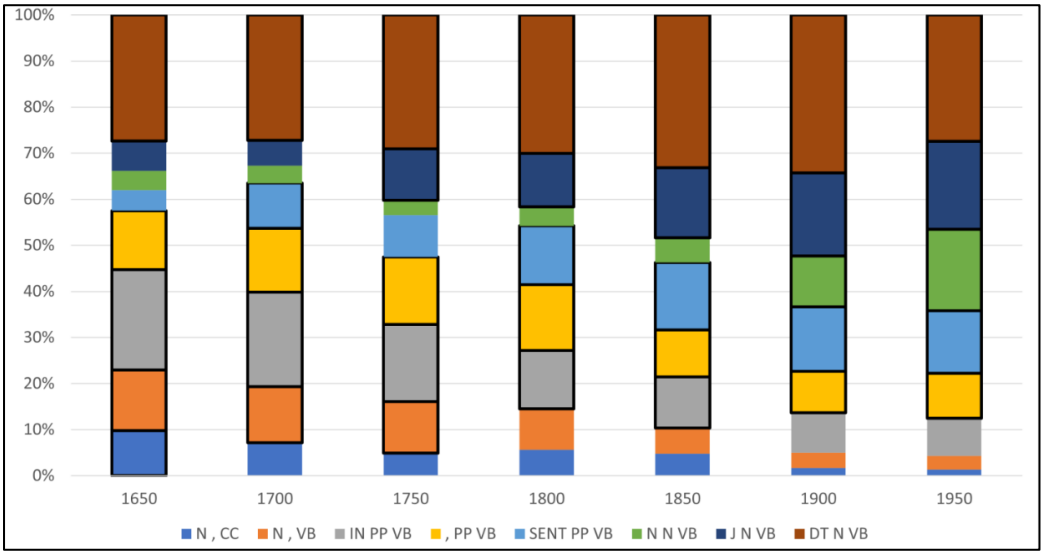

Figure 7. Development of trigrams preceding *not* + adjective in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

Since the surprisal of each word (cf. Section 5.2.1) is calculated given its preceding tokens, we check which lexical realizations contribute most to the decreasing surprisal of adjectives in the *not* + adjective pattern. To do so, we extract the five most frequent lexical trigrams preceding this pattern (Tab. 2). In the last time period (1950–1996), the most frequent lexical trigram (. *It is*) largely corresponds with the highly frequent preceding POS pattern [SENT-PP-VB], while the most frequent lexical trigrams in the first period are more heterogeneous (*"I thought it"*, *"that it was"*, *"that I was"*). Specifically, we can observe that in 1650–1700 not only the preceding trigrams but also the adjectives themselves are more varied (*amiss*, *possible*, *strange*, *able*), and not all of them have prefixed negated forms and/or could occur before nouns, while at the

ISSN 2453-8035

end of the 20th century all preceding contexts contain *it is / was* and three out of five adjectives are the same (*possible*).

Table 2. Top five lexical trigrams preceding *not* + adjective in RSC 6.0.1

1665–1699	1950–1996
I thought it not ***amiss***	*. It is not* ***possible***
that it was not ***possible***	*. It is not* ***clear***
that I was not ***able***	*, it is not* ***possible***
, it is not ***strange***	*. It is not* ***surprising***
, I am not ***able***	*. It was not* ***possible***

Therefore, we can confirm our assumption that the decrease in surprisal of adjectives negated by *not* derives from increasingly conventionalised contexts preceding such adjectives. In addition, we see that the contexts not only become more similar over time, but also less informationally dense, i.e., they contain overall more function words carrying less information (*it* instead of *I*, copula *be* instead of a lexical verb). Also, the trigrams in the last time period include a punctuation mark (generally carrying extremely low information content) pointing to a change in syntactic position of the *not* + adjective form from predominant use in subordinate clauses in 1665–1699 to the sentence-initial position in 1950–1996, i.e. functioning as an introductory, quite conventionalised evaluative phrase with low information content in the thematic position.

We further look at the right contexts of the construction by extracting the five most frequent subsequent POS trigrams. While the right context does not affect our surprisal values, it may however indicate the conditions under which the *not* + adjective form continues to be used. As shown in Fig. 8, there are three groups of trigrams following the analytic negation pattern: trigrams starting with *to*, e.g., [TO VV DT] (*to* – verb – determiner), trigrams starting with a preposition or subordinating conjunction, e.g., [IN DT N] (preposition / subordinating conjunction – determiner – noun) and a trigram starting with a comma [, IN DT] (comma – preposition / subordinating conjunction – determiner). The latter trigram represents cases where the adjective is not further defined, while the other two trigrams represent cases where the adjective is

ISSN 2453-8035

complemented by a *to*-infinitive or a prepositional phrase. In an aggregated version of Fig. 8, we summarise the trigrams into prepositional phrases (PP) and *to*-infinitives (TO INF) and the pattern starting with a comma (COMMA) (Fig. 9).

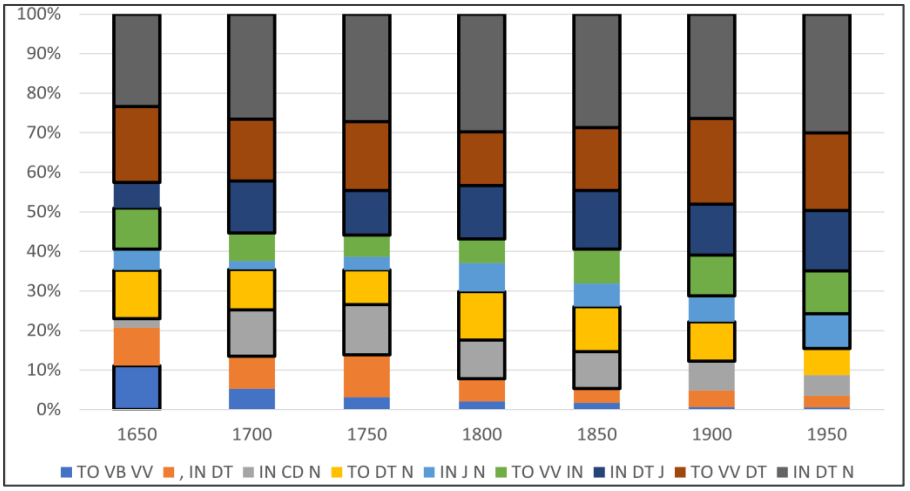

Figure 8. Development of trigrams following *not* + adjective in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

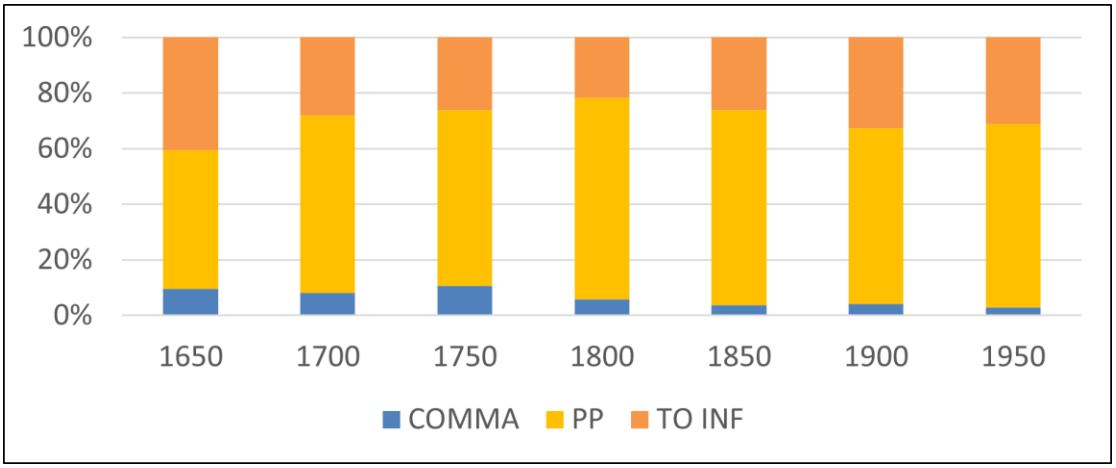

Figure 9. Five most frequent trigrams following *not* + adjective in RSC 6.0.1, aggregated into those followed by a comma and two tokens not further specified here (COMMA) indicating the occurrence of *not* + adjective in a clause-final position and those indicating the use of a *to*-infinitive clause (TO INF) or a prepositional phrase (PP) directly after the *not* + adjective sequence

We find that in all the time periods, analytically negated adjective constructions are most frequently followed by prepositional phrases (e.g., *not sensible of pain*). *To*-infinitives (e.g., *not possible to find*) represent the second most frequent pattern

ISSN 2453-8035

declining towards the mid-19th century and increasing again in the 20th century. The only pattern clearly fading out of the picture is the pattern starting with a comma, which represents negated adjectives without further defining information. We conclude from this that analytically negated adjectives continue to fulfil an important function in contexts where postmodification with a prepositional phrase or a *to*-infinitive is required. Postmodified adjectives cannot be placed in an attributive position, e.g., *"The person was not sensible of pain"* cannot be rephrased to *"*The non-sensible of pain person"* or *"the problem was not possible to be solved"* to *"*The not possible to solve problem."* A negated adjective that is not postmodified can more easily be prepended, e.g., *"The function was not linear" –> "the nonlinear function."* Note that *to*-infinitives can, for instance, be rephrased using affixation, encoding semantic information on the morphological level, i.e., "it was not possible to solve" can be rephrased as *"*It was *not solvable,"* or *"it was not easy to recognize"* as *"It was not easily recognizable,"* which is why they might be less frequent than prepositionally complemented adjectives.

Finally, considering a smaller window of context right of the *not* + adjective construction (the top ten unigrams following the analytic construction, Fig. 10) allows us to see in a more detailed way which grammatical contexts shape the use of *not* + adjective. We not only find the expected increase in the number of prepositions and subordinating conjunctions (IN) and the slightly decreasing proportion of *to*-infinitives (TO), but also an interesting proportional change in the overall sentence position of the analytically negated adjectives (sentence-final position [SENT] as compared to clause-final position [,]). As suggested by the trigram patterns, *not* + adjective increasingly occurs in the sentence-final position (around twice as often in 1950–1996 compared to 1665–1699), while its occurrence in the clause-final position (before a comma) declines by 50%.

 ISSN 2453-8035

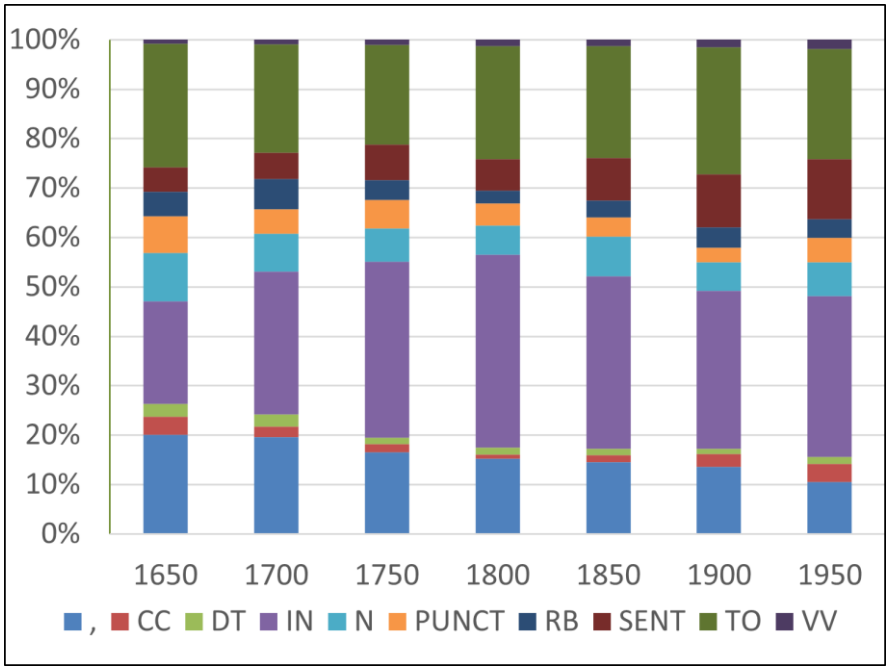

Figure 10. Top ten unigrams following *not* + adjective in RSC 6.0.1

In summary, analytically negated adjectives increasingly occur in conventionalised sentence-initial contexts, while the right context is dominated by prepositional phrases and *to*-infinitives. In terms of the sentence position, they seem to move to the beginning or end of a sentence, while their occurrence in the mid-sentence position seems to become less frequent over time.

5.2.2.2 Adjectives with prefix un-

Looking at the five most frequent POS trigrams preceding *un*-adjectives (Fig. 11), we find an increasing proportion of copular constructions including the verb *be* (VB), e.g., [N VB RB], [DT N VB] and [J N VB] where *un*-adjectives take the role of a predicative adjective phrase. Also, the most frequent pattern [N IN DT] (noun – preposition / subordinating conjunction – determiner) becomes more frequent. This trigram represents a context in which the adjective stands in the attributive position. Note that the context further to the left is potentially ambiguous since the POS tag IN can represent a preposition or a subordinating conjunction. The pattern may indicate the use of *un*-adjectives occurring in complex noun phrases of the type *"unequal distribution of the…", "undeniable proofs in the…"*) or of the type *"unexpected result*

 ISSN 2453-8035

that the…", "unequivocal evidence that the…"), although the latter is not frequently used in the RSC.

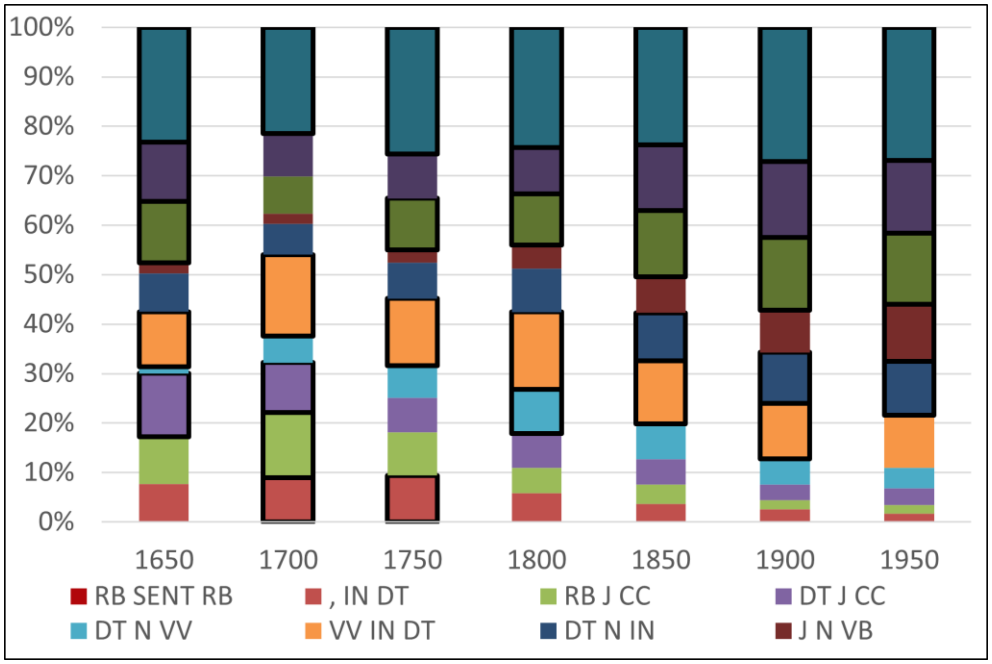

Figure 11. Development of trigrams preceding adjectives with *un-* in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

We aggregate the top 5 trigrams into three types of patterns: those that introduce the *un*-adjective in the attributive position (ATTRIBUTIVE), e.g., [DT J CC] (determiner – adjective – coordinating conjunction); those that introduce the adjective in the predicative position (PREDICATIVE), e.g., [DT N VB] (determiner – noun – verb *be*); those that are not categorizable into either of the two groups (NA) (see Fig. 12). We find that, indeed, *un*-adjectives increasingly occur in the predicative position in connection with verb phrases (40% in 1950–1996) while their use in the attributive position drops to slightly over 50% after 1850. These developments point to a versatility of usage options for *un*-adjectives. This versatility could explain the highest overall frequency of *un*-adjectives amongst the three patterns as well as the steep increase in the frequency of *un*-adjectives compared to *not* + adjective, which is confined to the predicative position.

ISSN 2453-8035

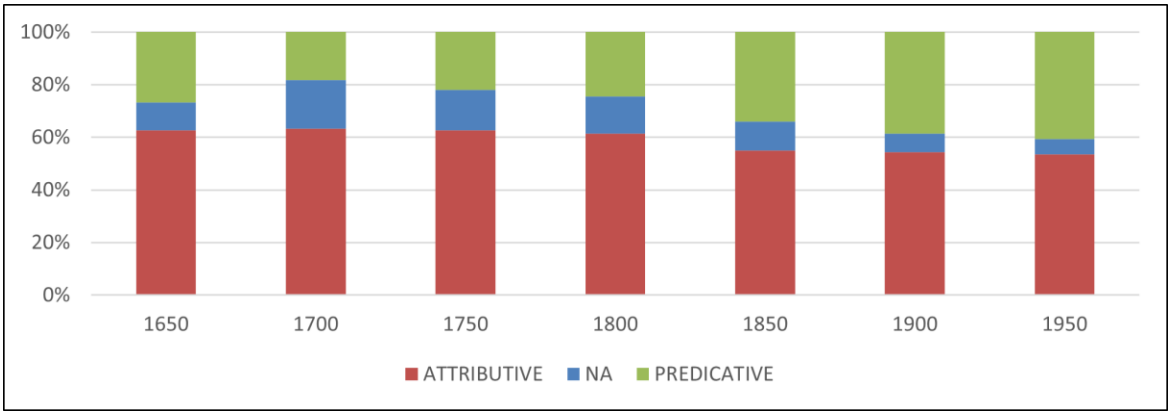

Figure 12. Five most frequent trigrams preceding adjectives with *un-* (cf. Fig. 11) in RSC 6.0.1, aggregated into predicative, attributive, and non-categorised (NA) patterns

Looking at the right contexts of *un*-adjectives (Fig. 13), we find a relatively stable distribution with three trigrams dominating the position throughout all the time periods: [N IN DT], [IN DT N], and [N IN N]. The three patterns confirm that *un*-adjectives most frequently occur immediately followed by a noun (attributive position) or followed by a preposition (predicative position). In addition, at the end of the 20th century *un*-adjectives in the attributive position occur increasingly at the end of sentences ([N SENT DT], [N) SENT]), while their occurrence in the clause-final position, e.g., [N , IN] (noun – comma – preposition or subordinating conjunction) decreases over time. This may point to a general shift of heavy and informationally dense noun phrases to the end of sentences.

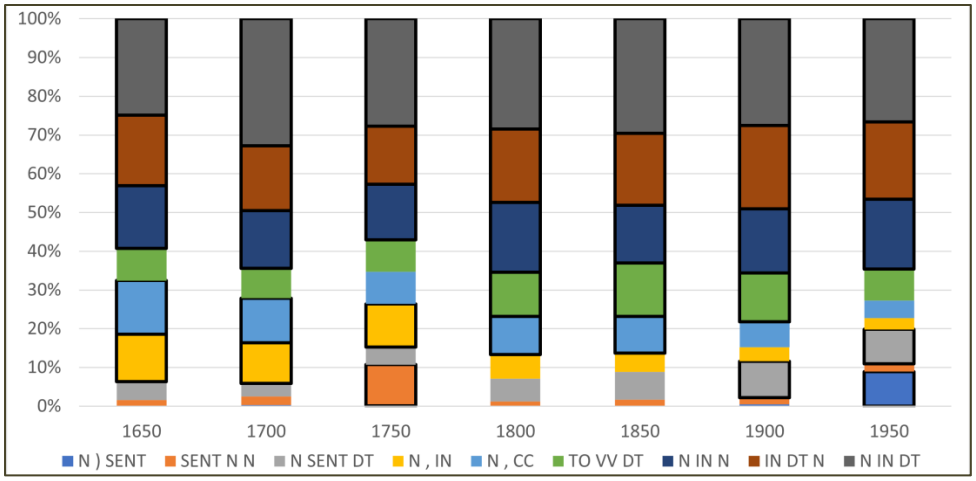

Figure 13. Development of trigrams following adjectives with *un-* in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

ISSN 2453-8035

A look into the most frequent lexical trigrams preceding *un*-adjectives reveals a similar development as the one found for the analytically negated adjectives (*not* + adjective) explaining the decline in surprisal. In the 17th century, the left lexical contexts of *un*-adjectives are lexically and grammatically heterogeneous, representing a mix of attributive and predicative positions of the *un*-adjectives. At the end of the 20th century, contexts are strongly conventionalised: four out of the five patterns represent sentence-initial (*". It is unlikely"*, *". It seems unlikely"*) and clause-initial (*", it is unlikely"*) copula patterns expressing evidentiality. Also, the *un*-adjectives in the first time period are more diverse (*unusual*, *unknown*, *unwilling*, *unequal*) than in the last period (*unlikely*, *unable*). Interestingly, the development of the lexical contexts of *un*-adjectives correlates strongly with that of the *not* + adjective pattern in that both increasingly occur in sentence-initial formulaic copula constructions of the type *". It is not* + adjective / *un*-adjective*"*.

Table 3. Top five lexical trigrams preceding adjectives with *un-* in RSC 6.0.1

1665–1699	1950–1996
, concerning an ***unusual***	*. It is* ***unlikely***
, with other ***unknown***	*, it is* ***unlikely***
which we are ***unwilling***	*. It seems* ***unlikely***
which was then ***unknown***	*, it seems* ***unlikely***
to be very ***unequal***	*we have been* ***unable***

5.2.2.3 Adjectives with prefix non-

A look into the five most frequent POS trigrams preceding *non*-adjectives (Fig. 14) indicates that in the 17th century the preceding contexts for *non*-adjectives were dominated by adjectives followed by coordinating conjunctions [DT J CC] (determiner – adjective – coordinating conjunction), [IN J CC] (preposition / subordinating conjunction – adjective – coordinating conjunction), [VB J CC] (verb *be* – adjective – coordinating conjunction). This shows that *non*-adjectives at earlier stages were mostly used in combination with other adjectives. In the following time periods, the coordinating conjunctions become less frequent, while trigrams representing parts of noun phrases preceding *non*-adjectives take over the slot, e.g., [N IN DT] (noun –

 ISSN 2453-8035

preposition or subordinating conjunction – determiner), [DT N IN] (determiner – noun – preposition or subordinating conjunction), [J N IN] (adjective – preposition or subordinating conjunction – determiner).

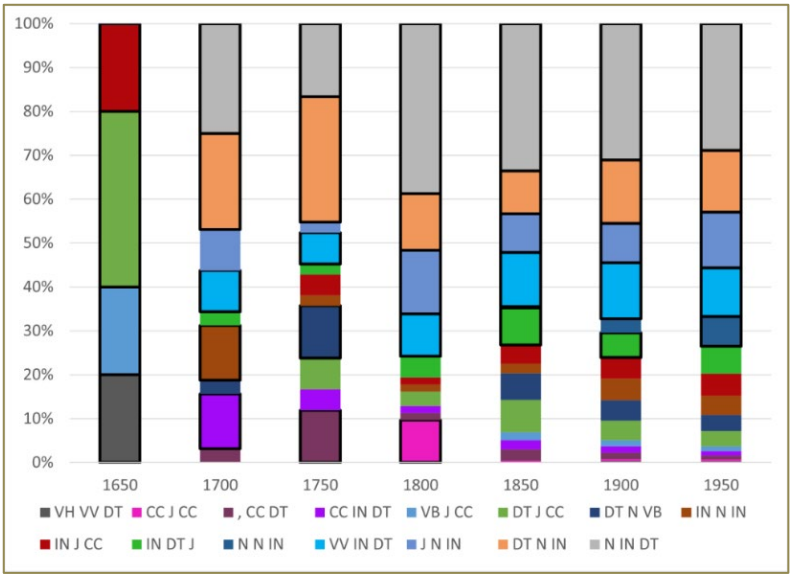

Figure 14. Development of trigrams preceding adjectives with *non-* in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

These contexts containing nouns and adjectives are informationally very dense, which could provide an explanation for the increasing surprisal values of the *non*-adjectives. Nouns usually carry more information than function words and make upcoming words less easily predictable.

Aggregating the entire group of the top five trigrams preceding the *non*-adjectives into predicative (patterns including the verb *be*) and attributive (those patterns that represent parts of noun phrases left of an attributive adjective) patterns (Fig. 15), we find a predominantly attributive use of *non*-adjectives throughout all the time periods. The right contexts (Fig. 16) are even less ambiguous since all of them start with a noun or an adjective followed by a noun. There is only one exception to this in the period between 1665 and 1699: [, RB VV] (comma – adverb – verb). These findings represent a contrast to adjectives negated with *un-*, which frequently occur both in attributive and predicative positions.

 ISSN 2453-8035

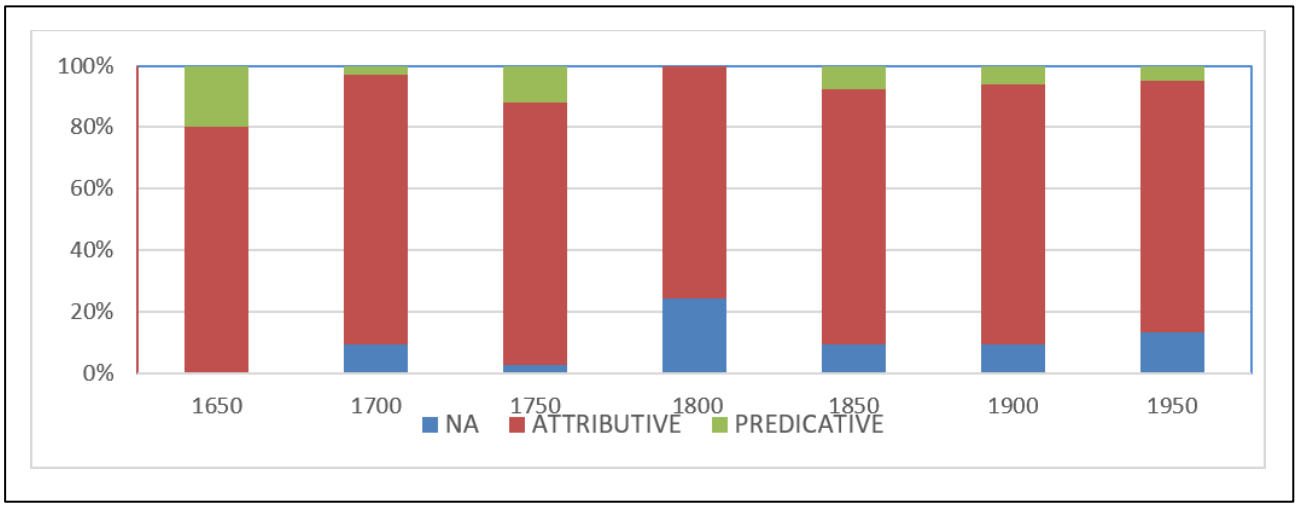

Figure 15. Five most frequent trigrams preceding adjectives with *non-* in RSC 6.0.1 aggregated into attributive, predicative and non-categorised (NA) patterns

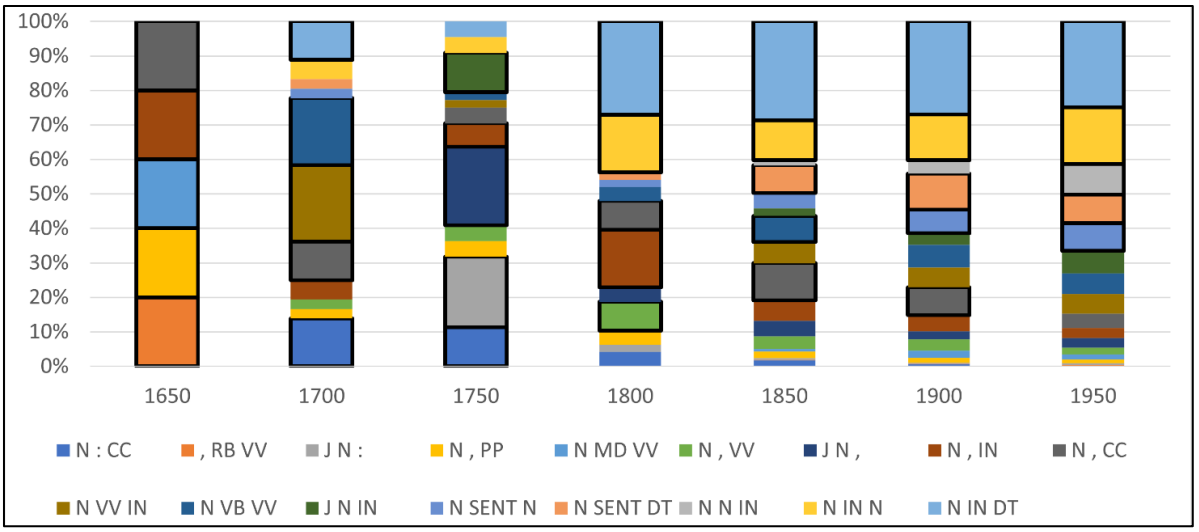

Figure 16. Development of trigrams following adjectives with *non-* in RSC 6.0.1 that were among the top five in at least one of the 50-year periods [the top 5 highlighted in the bar for each period respectively]

In 1665–1699, we see coordinate adjective structures as parts of noun phrases and predicative expressions among the top five POS trigrams of this time period (e.g., *"with impertinent and nonsensical"*, *"The new and non-descript"*, *"are new and nondescript"*, etc.) (Tab. 4). The prevailing *non*-adjective is *nondescript* (with and without hyphen). The most frequent lexical trigrams in 1950–1996 still contain coordinate adjectives (*linear and nonlinear*), but also parts of different types of complex noun phrases and prepositional phrases (predominantly with the recurrent adjective *nonlinear*). Some occur here again as frequent collocations as they are parts of constructions such as

ISSN 2453-8035

article titles that are quoted throughout different publications. We can assume that it is the more specific vocabulary in more complex noun phrases in contexts around *non-* adjectives that push surprisal values to increase in the later time periods.

Table 4. Top five lexical trigrams preceding adjectives with *non-* in RSC 6.0.1

1665–1699	1950–1996
with impertinent and ***nonsensical***	*1974 Linear and* ***nonlinear***
The new and ***non-descript***	*solutions of the* ***nonlinear***
such rare and ***non-descript***	*solution of the* ***nonlinear***
hath seen the ***nondescript***	*the linear and* ***nonlinear***
are new and ***nondescript***	*) On the* ***non-radial***

Another reason for the increase in surprisal may be a generally more diverse set of *non-* adjective types (Fig. 17).

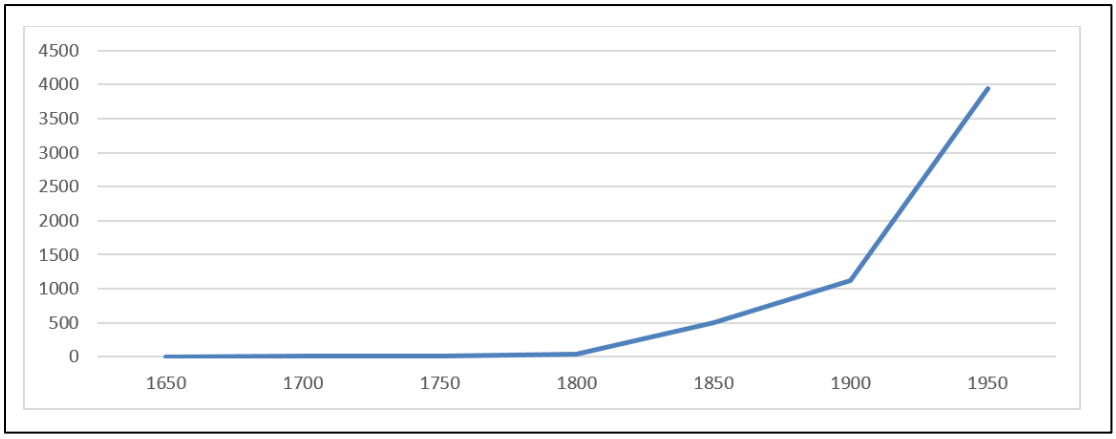

Figure 17. Different adjective types, (i.e. distinct word forms) with *non-* in the different time periods from the RSC 6.0.1

The fact that *non-* becomes increasingly frequent while *not* + adjective combinations become rare could suggest that *non-*adjectives gradually take over the function of predicative analytic adjectives in contexts where this is possible: e.g., *"the equation is not linear" – "the nonlinear equation"*. However, predicative adjectives with *to-* infinitives or complements cannot be moved to the attributive position (4 & 5):

(4) a) *The water was not proper for these Eels…* (RSC 6.0.1 text ID: 102744)

 ISSN 2453-8035

b) **The non-proper for these Eels water…*

(5) a) *… its spring is not sufficient to resist this pressure,…* (RSC 6.0.1 text ID: 102201)

b) * *… its non-sufficient to resist this pressure spring…*

5.2.2.4 Summary of micro-analysis

In summary, we have observed a diachronic change in use from analytic negation patterns towards synthetically negated adjectives in scientific journal articles. The analytically negated adjectives do not only become less frequent, but also less cognitively demanding over time as their predictability in context increases. Their decrease in information density is caused by strongly conventionalised left contexts, both on the grammatical (POS) and lexical levels. Grammatically, analytic negation patterns occur increasingly at the beginning of sentences, often embedded in copula constructions of the type ". *It is not* + adjective", or at the very end of sentences. Furthermore, the remaining adjectives negated with *not* increasingly occur in combination with a postmodifying *to*-infinitive or a prepositional phrase and represent cases that cannot easily be reformulated by using synthetically negated adjectives.

For both synthetically negated adjective types (*un-* and *non-*) we observe a strong increase in frequency between 1665 and 1996. While adjectives with *un-* are already very frequent in the 17th century, adjectives with *non-* soar in the last two centuries. The long-established adjectives with *un-* become increasingly predictable and less cognitively demanding over time settling in versatile, attributive, and increasingly predicative contexts. The predicative contexts are strongly conventionalised and possibly account for their decreasing surprisal overall. Like the analytically negated adjectives, adjectives with *un-* progressively gravitate towards the sentence beginning in copula constructions (*". It is un-*adjective*"*). Adjectives with *non-*, representing new members to the negation paradigm, represent the informationally most dense type of the three observed patterns. These adjectives seem to be confined to the attributive position increasingly occurring as parts of complex noun phrases such as *"solution of*

 ISSN 2453-8035

the non-linear equation".

6. Directions for future research

In the future, we would like to compare scientific language to other registers and to register-mixed diachronic corpora. For the comparative purposes related to diachronic aspects of synthetic and analytic adjective negation in the LModE period, we are planning to use the Corpus of Late Modern English Texts (CLMET, a relatively register-balanced corpus with a substantial part drawn from prose fiction, 1710–1920, De Smet et al. 2015) as a reference corpus. Scientific English and general English have become more distinct from one another over time (Degaetano-Ortlieb & Teich 2019). In the RSC, *un-* and *non-* become more frequent indicating a trend towards compressed negation patterns, but in the CLMET, we expect *un-* to become less frequent and adjectives prefixed with *non-* occurring only occasionally. As the variation between adjective negation via affixes and via analytic negation marker is not restricted to English, a contrastive study might be another interesting suggestion for future work, e.g., a comparison to German and French, where the status of *non(-)* or *nicht(-)* in front of adjectives is less clear than that of *non-* in English, so that these patterns are in between prefixed, morphological constructions and syntactic constructions with a negative particle (e.g., variants such as *non syndical* vs. *non-syndical* or *nichtabgeschlossen*, *nicht-abgeschlossen*, *nicht abgeschlossen* can be found, cf. Dugas 2014; Schneider 2020).

Another possible avenue for future research is the use of word embeddings of the RSC, a model that captures the usage patterns of the words and the distributionally most similar words in the corpus data (Teich et al. 2021). The interactive visualization (Word embeddings, *s.a.*) contains various clusters of multimorphemic adjectives that have formed and conventionalised particularly throughout LModE usage in scientific articles and that occur in close neighbourhood to each other. This application may provide useful insights on the diachronic development of synthetically negated forms

and structurally similar adjectives as well as their typical lexical contexts.

7. Conclusions

In the present paper we followed the hypothesis that prefixed negated adjectives become more frequent than analytically negated adjectives in scientific English over time. Our frequency-based analysis confirms the hypothesis showing that adjectives prefixed with *un-* and *non-* increase in frequency, while analytic forms of adjective negation become less frequent. In line with our second hypothesis, our surprisal-based analysis shows that analytically negated adjectives are less informationally dense (showing overall lower surprisal) than prefixed negated adjectives. By looking at the grammatical as well as lexical contexts of the negation patterns we found that analytically negated adjectives occur in increasingly conventionalised contexts and become therefore more easily predictable. Synthetic forms, while becoming more frequent, also become more diverse in terms of their contexts. Adjectives with the prefix *un-* are used versatilely in attributive and predicative positions. Adjectives with *non-* have become common premodifying items typically embedded in complex noun phrases carrying a high information load. Scientific language thus shows a trend towards informational densification due to using more condensed forms of adjective negation, while conserving less dense structures only where syntactically needed or where the formation of a synthetically negated adjective is blocked. Additionally, the function of prefixes concerning adding nuances and contributing to a finer adjustment of information than negation with *not* (as discussed in Section 2.1) seems to play an increasing role, at least with regard to the usage of *un-*.

Notes

1. "Affixal negation" / "morphological negation" / "synthetic negation" as well as "non-affixal negation" / "analytic negation" and the terms "patterns," "constructions" and "forms," respectively, are used synonymously in this paper.

2. In contrast to *un-*, *non-* has a hyphenated and a non-hyphenated form. The hyphenated form is the dominant one throughout the data. To a certain extent, the

ISSN 2453-8035

existence of individual spelling variants for some forms might also have an influence on surprisal.

Abbreviations

BNC – British National Corpus

CQP – Corpus Query Processor

CLMET – Corpus of Late Modern English texts

EModE – Early Modern English

LModE – Late Modern English

OCR – Optical character recognition

OED – Oxford English Dictionary

PDE – Present-Day English

POS – part-of-speech (individual part-of-speech name abbreviation from the tagset used in this paper can be found in the Penn Treebank tagset, *s.a.*)

RSC – Royal Society Corpus

References

Aina, L., Bernardi, R. & Fernández, R. (2018). A distributional study of negated adjectives and antonyms. In *Proceedings of the 5th Italian conference on computational linguistics, CLiC-it 2018 (December, 2018, Torino)*. Cabrio, E., Mazzei, A. & Tamburini, F. Available at: http://ceur-ws.org/Vol-2253/

Allen, M. (1978). The morphology of negative prefixes in English. In *North East linguistics society*, 8 (1), p. 1-11.

Baayen, R.H. & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. In *Linguistics*, 29 (5), p. 801-843. DOI: 10.1515/ling.1991.29.5.801.

Baldi, P., Broderick, V. & Palermo, D.S. (1985). Prefixal negation of English adjectives: Psycholinguistic dimensions of productivity. In *Historical semantics – Historical word-formation.* Fisiak, J. (ed.). The Hague: Mouton, p. 33-58.

Bauer, L. (2001). *Morphological productivity.* Cambridge: CUP.

Bauer, L., Lieber, R. & Plag, I. (2013). *The Oxford reference guide to English morphology.* Oxford: OUP.

Bauer, L. & Nation, I.S.P. (2020). *English morphology for the language teaching profession.* New York: Routledge.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison.* Cambridge: CUP.

Biber, D. & Finegan, E. (1997). Diachronic relations among speech-based and written registers in English. In *To explain the present: Studies in the changing English language in honour of Matti Rissanen.* Nevalainen, T. & Kahlas-Tarkka, L. (eds.). Helsinki: Société Néophilologique, p. 253-275.

Biber, D. & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. In *Journal of English for academic purposes*, 9 (1), p. 2-20.

Biber, D. & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing.* Cambridge: CUP.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow – Essex: Pearson Education.

Blanco, E. & Moldovan, D. (2011). Some issues on detecting negation from text. In *Proceedings of the 24th international Florida artificial intelligence research society conference FLAIRS - 24*, Palm Beach, FL, USA, May 18-20, 2011, p. 228-233.

Calude, A.S. & Bauer, L. (2021). *Mysteries of English grammar – A guide to complexities of the English language.* Abingdon – New York: Taylor & Francis.

Cartoni, B. & Lefer, M.-A. (2011). Negation and lexical morphology across languages: Insights from a trilingual translation corpus. In *Poznan studies in contemporary linguistics*, 47 (4), p. 795-843.

Charles, W.G. & Miller, G.A. (1989). Contexts of antonymous adjectives. In *Applied psycholinguistics*, 10 (3), p. 357-375.

Croft, W. (1991). The evolution of negation. In *Journal of linguistics*, 27 (1), p. 1-27.

De Clercq, K. & Vanden Wyngaerd, G. (2019). Negation and the functional sequence.

In *Natural language & linguistic theory*, 37 (2), p. 425-460.

Degaetano-Ortlieb, S., Kermes, H., Khamis, A. & Teich, E. (2019). An information-theoretic approach to modeling diachronic change in scientific English. In *From data to evidence in English language research.* Suhr, C., Nevalainen, T. & Taavitsainen, I. (eds.). Leiden – Boston: Brill, p. 258-281.

Degaetano-Ortlieb, S. & Teich, E. (2016). Information-based modeling of diachronic linguistic change: From typicality to productivity. In *Proceedings of the 10th SIGHUM workshop on language technology for cultural heritage, social sciences and humanities (LaTeCH)*. Reiter, N., Alex, B. & Zervanou, K.A. (eds.). Berlin: Association for Computational Linguistics, p. 165-173.

Degaetano-Ortlieb, S. & Teich, E. (2019). Toward an optimal code for communication: the case of scientific English. In *Corpus linguistics and linguistic theory* [online publication ahead of print]. DOI: 10.1515/cllt-2018-0088. Available at: https://www.degruyter.com/document/doi/10.1515/cllt-2018-0088/html

De Smet, H., Flach, S., Tyrkkö, J. & Diller, H.-J. (2015). *The corpus of Late Modern English texts (CLMET), Version 3.1: Improved tokenization and linguistic annotation.* Available at: https://hdl.handle.net/21.11119/0000-0002-43F3-0

Dixon, R.M.W. (2014). *Making new words: Morphological derivation in English.* Oxford: OUP.

Dugas, E. (2014). [non-Adj] sequences in French – Morphological negation, syntactic negation, or in-between? In *Morphology and its interfaces: Syntax, semantics and the lexicon.* Amiot, D., Tribout, D., Grabar, N., Patin, C. & Tayalati, F. (eds.) [Lingvisticæ Investigationes, 37 (2)], p. 194-208.

Farshchi, S., Andersson, A., van de Weijer, J. & Paradis, C. (2020). Processing sentences with sentential and prefixal negation: An event-related potential study. In *Language, cognition and neuroscience*, 36 (1), p. 84-98.

Fischer, S., Knappen, J., Menzel, K. & Teich, E. (2020). The Royal Society Corpus 6.0. Providing 300+ years of scientific writing for humanistic study. In *Proceedings of the 12th international conference on language resources and evaluation, LREC'20,*

Marseille, France, May 13-15, 2020, p. 794-802.
Funk, W.-P. (1971). Adjectives with negative affixes in Modern English and the problem of synonymy. In *Zeitschrift für Anglistik und Amerikanistik*, 19 (4), p. 364-386.
Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium.* Marantz, A., Miyashita, Y. & O'Neil, W. (eds.). Cambridge, MA – London: MIT Press, p. 95-126.
Gilquin, G. & Lefer, M.-A. (2017). Exploring word-formation in learner corpus research: A case study on English negative affixes. Paper presented at the *4th Learner corpus research conference*, Bolzano, Italy, October 5-7, 2017. Abstract available at: https://lcr2017.eurac.edu/
Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd meeting of the North American chapter of the association for computational linguistics on language technologies*, Pittsburgh, Pennsylvania, USA, June 2-7, 2001, p. 1-8.
Halliday, M.A.K. (1985). *An introduction to functional grammar.* 1st ed. London: Edward Arnold.
Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. In *International journal of corpus linguistics*, 17 (3), p. 380-409.
Horn, L.R. (2001). *A natural history of negation.* Stanford: CSLI.
Hulse, V. (2010). *Productivity in morphological negation: A corpus-based approach.* PhD thesis. Manchester: University of Manchester.
Jackendoff, R.S. (1969). An interpretive theory of negation. In *Foundations of language*, 5 (2), p. 218-241.
Kastovsky, D. (2006). Vocabulary. In *A history of the English language*. Hogg, R.M. & Denison, D. (eds.). Cambridge: Cambridge University Press, p. 199-270.
Kermes, H., Degaetano-Ortlieb, S., Khamis, A., Knappen, J. & Teich, E. (2016). The Royal Society Corpus: From uncharted data to corpus. In *Proceedings of the 10th international conference on language resources and evaluation (LREC'16)*, Portorož,

Slovenia, May, 23-28, 2016, p. 1928-1931.
Juzek, T., Krielke, M.-P. & Teich, E. (2020). Exploring diachronic syntactic shifts with dependency length: The case of scientific English. In *Proceedings of the 4th workshop on universal dependencies (UDW 2020)*. De Marneffe, M.-C., de Lhoneux, M., Nivre, J. & Schuster, S. (eds.) Barcelona: Association for Computational Linguistics, p. 109-119.
Jespersen, O. (1917). *Negation in English and other languages.* Copenhagen: A.F. Høst.
Jiménez-Zafra, S.M., Morante, R., Martín-Valdivia, M.T. & Ureña-López, L.A. (2020). Corpora annotated with negation. In *Computational linguistics*, 46 (1), p. 189-244.
Jones, S. (2002). *Antonymy: A corpus-based perspective.* London: Routledge.
Joshi, S. (2020). Affixal negation. In *The Oxford handbook of negation.* Déprez, V. & Espinal, M.T. (eds.). Oxford: OUP, p. 75-88.
Joshi, S. (2012). Affixal negation – direct, indirect and their subtypes. In *Syntaxe et sémantique*, 13 (1), p. 49-63.
Kjellmer, G. (2005). Negated adjectives in Modern English. A corpus-based study. In *Studia neophilologica*, 77 (2), p. 156-170.
Kwon, H.-S. (1997). Negative prefixation from 1300 to 1800: A case study in in-/un- variation. In *ICAME Journal*, 21, p. 21-42.
Kwon, H.-S. (2001). Using corpora for the study of word formation: A case study in English negative prefixation. In *Korean journal of English language and linguistics*, 1 (3), p. 369-386.
Lefer, M.-A. (2012). Word-formation in translated language: The impact of language-pair specific features and genre variation. In *Across languages and cultures*, 13 (2), p. 145-172.
Levy, R.P. (2008). Expectation-based syntactic comprehension. In *Cognition*, 106 (3), p. 1126-1177.
Lieber, R. (2004). *Morphology and lexical semantics.* Cambridge: CUP.
Marchand, H. (1969). *The categories and types of present-day English word-formation*. 2nd ed. Munich: Verlag C.H. Beck.
Mazzon, G. (2004). *A history of English negation.* London: Pearson Longman.
Menzel, K. (2021). Scientific eponyms throughout the history of English scholarly

journal articles. In *Broadening perspectives in the history of dictionaries and word studies.* Van de Velde, H. & Dolezal, F.T. (eds.). Newcastle upon Tyne: Cambridge Scholars Publishing, p. 159-193.

Menzel, K. & Degaetano-Ortlieb, S. (2017). The diachronic development of combining forms in scientific writing. In *Lege artis. Language yesterday, today, tomorrow. The journal of University of SS Cyril and Methodius in Trnava.* Warsaw: De Gruyter Open, II (2), December 2017, p. 185-249. DOI: 10.1515/lart-2017-0016

Menzel, K., Knappen, J. & Teich, E. (2021). Generating linguistically relevant metadata for the Royal Society Corpus. In *Challenges in combining structured and unstructured data in corpus development – special issue of research in corpus linguistics, Asociación Española de Lingüística de Corpus (AELINCO)*, 9 (1), p. 1-18.

Murphy, G.L. & Andrew, J.M. (1993). The conceptual basis of antonymy and synonymy in adjectives. In *Journal of memory and language*, 32 (3), p. 301-319.

Okada, A. (2010). The analyses of the English negative prefixes in the history of English. In *Aspects of the history of English language and literature.* Imahayashi, O., Nakao, Y. & Ogura, M. (eds.). Frankfurt: Peter Lang, p. 343-351.

Orwell, G. (1954): *1984: A novel.* Harmondsworth: Penguin Books.

Oxford English dictionary. (1884–1928, 1st ed. ('A new English dictionary on historical principles') / 1989, 2nd ed. / 2000, 3rd ed.). Oxford: Clarendon Press. 3rd ed. Available at: http://www.oed.com/

Oxford English dictionary online, non-, prefix. Available at: https://www.oed.com/view/Entry/127776

Oxford English dictionary online, un-, prefix[1]. Available at: https://www.oed.com/view/Entry/208915

Penn Treebank tagset. (s.a.). Available at: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Royal Society Corpus (RSC 6.0 Open – open release version of texts until 1920,

available at: https://hdl.handle.net/21.11119/0000-0004-8E37-F, RSC 6.0.1 full corpus version with texts until 1996, available upon request via Saarbrücken CQPweb, available at: http://corpora.clarin-d.uni-saarland.de/cqpweb/)

Russell, B. (1948). *Human knowledge: Its scope and limits.* New York: Simon & Schuster.

Santorini, B. (1990). *Part-of-speech tagging guidelines for the Penn treebank project.* [Technical report]. 3rd revision. University of Pennsylvania. Available at: https://repository.upenn.edu/cis_reports/570/

Schneider, R. (2020). Unumstritten oder nicht beantwortbar – Eine Korpusstudie zur Variation in der Negation von Adjektiven im Deutschen. In *Regensburg papers in linguistics*, 15. Regensburg: Universitätsbibliothek Regensburg. Available at: https://epub.uni-regensburg.de/41419/1/RPiL 15_Schneider.pdf

Saarbrücken CQPweb. Available at: http://corpora.clarin-d.uni-saarland.de/cqpweb/

Shannon, C.E. (1948). A mathematical theory of communication. In *Bell system technical journal*, 27 (3), p. 379-423.

Smitterberg, E. (2021). *Syntactic change in Late Modern English: Studies on colloquialization and densification.* Cambridge: CUP.

Štekauer, P. (2000). *English word-formation: A history of research, 1960–1995.* Tübingen: Gunter Narr Verlag.

Teich, E., Fankhauser, P., Degaetano-Ortlieb, S. & Bizzoni, Y. (2021). Less is more/more diverse: On the communicative utility of linguistic conventionalization. In *Frontiers in communication*, 5, Article 620275, p. 1-17.

Tottie, G. (1999). Affixal and non-affixal negation – A case of stable variation over time? In *Negation in the history of English.* Tieken-Boon van Ostade, I., Tottie, G. & van der Wurff, W. (eds.). Berlin: Mouton de Gruyter, p. 233-265.

Tottie, G. (1980). Affixal and non-affixal negation in English – two systems in (almost) complementary distribution. In *Studia linguistica*, 34 (2), p.101-123.

Tottie, G. (1991). *Negation in English speech and writing.* San Diego – New York – London: Academic Press.

Word embeddings of the RSC. [Description, models, and visualization]. Available at:

http://corpora.ids-mannheim.de/openlab/diaviz1/description.html & http://corpora.ids-mannheim.de/openlab/diaviz1/flying-bubbles.html#embeddings=_rsc-diachron-1929-perplexity50-init-tc0-t1

Van Son, C., van Miltenburg, E. & Morante, R. (2016). Building a dictionary of affixal negations. In *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics (ExProM 2016),* Osaka, Japan, December 12, 2016, p. 49-56.

Zangwill, N. (2011). Negative properties. In *Noûs*, 45 (3), p. 528-556.

Zimmer, K.E. (1964). Affixal negation in English and other languages: An investigation of restricted productivity. In *Supplement to word*, 20 (2).

Contact data

Author #1

name:	**Katrin Menzel**
academic title / rank:	Dr Phil. (Language and translation studies) Post-doctoral researcher
department:	Department of Language Science and Technology
institution:	Saarland University, Campus A2.2, 66123 Saarbrücken, Germany
e-mail:	k.menzel@mx.uni-saarland.de
fields of interest:	English historical linguistics, contrastive linguistics, corpus linguistics, morphology, language contact, translatology.

Author #2

name:	**Marie-Pauline Krielke**
academic title / rank:	MA (Conference interpreting) Doctoral researcher
department:	Department of Language Science and Technology
institution:	Saarland University, Campus A2.2, 66123 Saarbrücken, Germany
e-mail:	mariepauline.krielke@uni-saarland.de
fields of interest:	Diachronic register variation, grammatical change, English linguistics, contrastive linguistics, corpus linguistics, discourse studies, translatology.

 ISSN 2453-8035

Author #3

name:	**Stefania Degaetano-Ortlieb**
academic title / rank:	Priv.-Doz. (English linguistics and corpus linguistics), Dr Phil. (Applied linguistics) Senior lecturer / Assistant Professor
department:	Department of Language Science and Technology
institution:	Saarland University, Campus A2.2, 66123 Saarbrücken, Germany
e-mail:	s.degaetano@mx.uni-saarland.de
fields of interest:	Text mining, data analytics, sociolinguistics, register/language variation, diachronic change in language use, English linguistics.

Résumé

This paper examines the development of synthetic and analytic adjective negation in an English diachronic corpus of scientific journal articles (Royal Society Corpus 6.0.1 – RSC, 1665–1996). We specifically focus on synthetically negated adjectives with the prefix *un-* or *non-* (e.g., *unavoidable*, *non-magnetic*) and analytic patterns in adjectives after the negation marker *not* (e.g., *not avoidable*, *not magnetic*). The use of more condensed synthetic forms contributes to economy of expression and represents a register-specific strategy used by writers to optimise the information density distribution in linguistic structures with regard to the expectations and processing capacity of their addressees. Synthetically negated English adjectives contribute to word-internal compression as multimorphemic noun premodifiers or predicative expressions (e.g., *"non-magnetic metals", "unavoidable difficulties"*) in contrast to clausal structures containing positive adjectives and the analytic negation marker *not* (e.g., in *"metals which are not magnetic", "difficulties are not avoidable"*), which often represent less compressed variants. Besides comparing frequency distributions over time, we use surprisal from an information-theoretic framework to measure the amount of information that particular patterns carry. Overall, we find that analytic patterns of adjective negation become less frequent in scientific writing, but more conventionalised – and therefore more easily predictable – in their textual contexts.

 ISSN 2453-8035

Prefixed negated adjectives become more frequent and more diverse with regard to their contexts. Adjectives with the negation marker *un-* are used in a more versatile way in attributive and predicative positions. Adjectives with *non-*, an initially rare word formation pattern, become established as common lexemes in scientific terminology, typically embedded as attributive modifiers in complex noun phrases carrying a high information load. English scientific journal articles thus show a trend towards informational densification by using more condensed forms of adjective negation while conserving less dense structures primarily in contexts where they are syntactically or grammatically required.

Key words: adjective negation, prefixation, clausal negation, scientific English, corpus-based diachronic analysis, surprisal.

Article was received by the editorial board 14.01.2022;
Reviewed 26.03.2022 and 26.03.2022.
Similarity Index 3%