

What does the adjective *novel* mean now? Collocational change in biomedical research discourse

Neil Millar¹, Bojan Batalo²

¹University of Tsukuba, ²National Institute of Advanced Industrial Science and Technology (AIST)
¹1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan, ²2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan
millar.neil@u.tsukuba.ac.jp, bojan.batalo@aist.go.jp

Abstract

Collocation is often used to investigate lexical meaning, yet comparatively little attention has been paid to the analysis of collocational behaviour over time. This paper explores a framework for diachronic dependency-based collocation analysis through a case study of the adjective *novel* in biomedical research discourse. Using a corpus of 901,717 National Institutes of Health (NIH) funded research project abstracts (355.8 million words) spanning 1985–2020, we examine changes in the dependency-based collocations of *novel*. We explore two descriptive measures that capture different aspects of collocational behaviour: the proportion of occurrences accounted for by individual collocates (Collocate Prominence) and the degree to which collocates preferentially occur with *novel* relative to their overall frequency (Collocate Preference). The analyses indicate a broadening in the collocational profile of *novel*. The adjective increasingly modifies both abstract concepts and emerging biomedical entities, while coordinated adjectives and modifying adverbs indicate growing associations with evaluative and promotional language. We evaluate the strengths and limitations of the proposed methodology and discuss the findings in relation to usage-based accounts of semantic change and broader shifts in the rhetorical culture of biomedical research.

Keywords: diachronic collocation, semantic change, dependency-based collocation, biomedical discourse

1. Introduction

The adjective *novel* has become a routine descriptor in contemporary scientific discourse, particularly in biomedical research. Its core meaning denotes something new, original, or previously unknown. Yet its frequency has risen at a rate that is difficult to explain solely by growth in scientific activity. In abstracts describing projects funded by the U.S. National Institutes of Health (NIH), *novel* appeared in 9% of project abstracts in 1985 but in 36% by 2020. Its normalised frequency increased from 147 to 1,201 words per million, an increase of 717% (Millar et al., 2022). This trend raises the question of whether the meaning of *novel* has changed over time and, if so, how.

Collocation provides a way of investigating lexical meaning because recurrent associations form part of what speakers know about a word (Firth, 1957; Sinclair, 1991; Stubbs, 2001). Lexical Priming offers an account of how such associations develop: repeated encounters with a word in particular linguistic, discourse, and social contexts prime speakers to expect particular collocations (Hoey, 2005). This view is consistent with usage-based approaches, which regard linguistic structure as emerging from language use (Bybee, 2010, 2015). Thus, changes in patterns of use, including collocation, should provide evidence of semantic change. While collocation is widely used in synchronic corpus linguistics, comparatively little attention has been paid to how collocational behaviour can be analysed over time.

This paper explores a framework for investigating diachronic collocational change through a case study of *novel* in biomedical research discourse. Using a corpus of National Institutes of Health (NIH) funded research abstracts spanning 1985–2020, we examine changes in the dependency-based collocations of *novel*. Rather than identifying collocates on the basis of a fixed span, we focus on direct dependency relations that identify the nouns described as *novel*, the adjectives coordinated with *novel*, and the adverbs that modify *novel*. To characterise changes over time, we explore two descriptive measures: (1) Collocate Prominence, the proportion of occurrences of *novel* accounted for by a particular collocate, and (2)

Collocate Preference, the degree to which a collocate preferentially occurs with *novel* relative to its overall frequency in the corpus.

The study addresses three questions. First, has the collocational profile of *novel* broadened over time? Second, what changes are observed in the noun collocates of *novel*, particularly with respect to abstract concepts and emerging biomedical entities? Third, has *novel* become increasingly associated with evaluative and promotional language through its adjective and adverb collocates?

2. Background and motivation

2.1 Semantic change and collocational behaviour

Usage-based approaches view semantic change as emerging through the interaction of language use and domain-general cognitive processes (Bybee, 2010, 2015). As words are repeatedly used in *new* contexts, speakers both produce and encounter these *new* uses. Through processes such as categorization and analogy, these repeated experiences reinforce *new* patterns of use until they become increasingly conventional or entrenched. For example, *awesome*, once largely reserved for experiences that inspired awe, is now routinely applied to everyday objects, events and experiences. Semantic change is therefore viewed as a gradual process of extension and entrenchment rather than the abrupt replacement of one meaning by another.

One consequence of this dynamic is that semantic change is typically gradual and distributional. Older and newer uses often coexist, but differ in the linguistic environments in which they occur and in their relative frequencies. If the meaning of *novel* has changed over time, we would therefore expect to observe changes not only in how frequently it is used, but also in the kinds of concepts it characterises, the adjectives with which it is coordinated, and the adverbs that modify it. Because these contextual shifts are realised linguistically, they can, we speculate, be investigated through changes in collocational behaviour.

Historical semantic research has identified a number of pathways through which lexical meanings develop. Of relevance to this study, semantic broadening extends a word to a wider range of referents or situations. Hyperbolic

extension occurs when a strongly evaluative expression is applied to referents or situations that do not fully warrant its original force. As hyperbolic use becomes increasingly common in comparatively ordinary contexts, the exaggeration may become conventional and the expression may lose some of its earlier semantic force. This process, often described as semantic bleaching, results in a word becoming less semantically specific or less evaluatively forceful while remaining appropriate across an increasingly diverse range of contexts. Although these developments arise through different cognitive and pragmatic mechanisms, they all predict systematic changes in the contexts in which words are used.

These developments should therefore leave different traces in collocational behaviour. Broadening would be expected to increase the diversity of concepts described as *novel*. Hyperbolic extension and semantic bleaching, by contrast, are more likely to be reflected in the evaluative phraseology surrounding the adjective, including the adjectives with which it is coordinated and the adverbs that modify it.

2.2 Interpreting diachronic collocational change

Collocational analyses have generally been applied synchronically. An exception is Hilpert's (2006, 2012) work on diachronic collocational analysis, which examines changes in the lexical preferences of grammatical constructions across extended historical periods. The present study addresses a related but distinct problem: tracing changes in the dependency-based collocations of a lexical item across 36 annual time points.

Extending collocational analyses diachronically presents methodological challenges. Frequency can generally be compared directly once corpus size has been normalised. Collocational behaviour, however, is inherently multidimensional because it reflects not only how often a word occurs but also the linguistic environments in which it occurs. From a diachronic perspective, this amounts to comparing two moving targets: the frequency of the node changes over time, but so too do the frequencies and distributions of its collocates. Consequently, observed changes in collocational behaviour may reflect changes in the node, the collocates, or both simultaneously.

A further consideration concerns how diachronic collocational change should be characterised. Corpus linguistics has developed numerous association measures, each reflecting different aspects of the relationship between a node and its collocates (Evert, 2009). No single measure is therefore appropriate for all analytical purposes. More recently, Gries (2024) has argued that commonly used association measures are confounded with frequency and that corpus linguistics should move away from describing complex linguistic phenomena using a single statistic. Instead, he advocates a "tupleization" approach in which different dimensions of information are retained separately for description and interpretation.

The present study adopts a similar philosophy, although for a different purpose. Our objective is not to identify the statistically strongest collocations of *novel* at individual points in time, but to describe how its collocational profile evolves over time. We therefore use two simple descriptive proportions rather than a single association score. Conceptually, these correspond to the two directions of conditional probability: one considers, given an occurrence of *novel*, how frequently a particular collocate occurs with it, while the other considers, given a particular collocate,

how strongly it is associated with *novel* relative to its overall frequency in the corpus. Their formal definitions are presented in the following section.

2.3 Biomedical research discourse as a case study

Biomedical research provides a useful setting for investigating diachronic semantic change. Scientific output has expanded substantially, competition for funding and publication has intensified, and studies have documented increasing use of promotional and evaluative language in scientific writing, particularly around novelty, importance, certainty, and potential impact (Fraser & Martin, 2009; Vinkers et al., 2015; Hyland & Jiang, 2016, 2021).

The present study examines abstracts from all projects funded by the U.S. National Institutes of Health (NIH). The corpus comprises 901,717 project abstracts spanning 1985–2020, providing 36 annual observations. Previous analyses have documented diachronic shifts in promotional language and epistemic stance in this corpus (Millar et al., 2022; Millar et al., 2023).

Among these changes, *novel* increased from 147 to 1,201 words per million between 1985 and 2020, a rise of 717% (Millar et al., 2022). By comparison, *new* increased from 1,017 to 1,638 words per million, or 61%. This substantially faster growth suggests that the rise of *novel* cannot be explained simply by increased reference to newness. The present study therefore examines whether its increased frequency was accompanied by systematic changes in collocational behaviour.

3. Method

3.1 Corpus

The corpus comprised abstracts describing projects funded by the U.S. National Institutes of Health (NIH) and related agencies between 1985 and 2020. Records were downloaded from the NIH ExPORTER database (<https://exporter.nih.gov/>) and, after removing duplicates and records without abstracts, the final corpus contained 901,717 project abstracts (355.8 million words) spanning 36 annual time points. The corpus was automatically annotated for part-of-speech using the Stanza NLP toolkit (Qi et al., 2020), indexed in CQPweb (Hardie, 2012), and queried for occurrences of *novel*. This yielded 329,007 occurrences in 320,354 sentences, which formed the dataset for the dependency-based analyses.

3.2 Dependency-based collocation extraction

All sentences containing *novel* were exported from CQPweb and processed using spaCy (Honnibal et al., 2020) together with custom Python scripts. We extracted words participating in four syntactic relations with *novel*: (i) nouns modified by *novel* in attributive constructions (e.g. a *novel* method), (ii) nouns linked to *novel* in predicative constructions (e.g. this method is *novel*), (iii) adjectives coordinated with *novel* (e.g. *novel* and innovative), and (iv) adverbs modifying *novel* (e.g. highly *novel*).

To assess extraction accuracy, 595 sentences were manually annotated from five samples distributed across the extracted dataset of sentences containing *novel*, including both attributive and predicative constructions. Across these evaluation sets, recall ranged from 96–99% and precision from 94.5–100%.

3.3 Measures of diachronic collocational change

To characterise changes in collocational behaviour over time, we calculated two complementary descriptive measures. The measures capture two different conditional perspectives on the relationship between *novel* and its collocates and correspond broadly to the forward and reverse directions of conditional probability.

Collocate Prominence measures the proportion of occurrences of *novel* accounted for by a particular collocate:

$$Prominence(c) = \frac{f(novel, c)}{f(novel)}$$

where $f(novel, c)$ is the frequency with which *novel* occurs with collocate c , and $f(novel)$ is the total frequency of *novel* in that year.

Collocate Preference measures the extent to which a collocate preferentially occurs with *novel* relative to its overall frequency in the corpus:

$$Preference(c) = \frac{f(novel, c)}{f(c)}$$

where $f(c)$ is the total frequency of the collocate in the corpus for that year.

3.4 Diachronic analysis

Values of Collocate Prominence and Collocate Preference were calculated separately for each year from 1985 to 2020. Trends were assessed using Kendall's rank correlation coefficient (τ), following Hilpert and Gries (2009).

Changes in the two measures were interpreted jointly. Existing collocates may increase in both prominence and preference, indicating that they account for a growing proportion of occurrences of *novel* while also becoming more strongly associated with it. Alternatively, preference may increase while prominence remains stable or decline as *novel* expands into a wider range of collocational contexts. Diachronic analyses may also indicate the emergence of previously absent or increasingly prominent collocates, reflecting changes in the concepts with which the adjective is associated.

Because semantic broadening was hypothesised to involve increasing associations with abstract concepts, noun collocates were classified according to whether they were included in Schmid's (2000) inventory of 670 shell nouns - nouns referring to abstract conceptual entities such as idea, possibility, and evidence. The remaining noun collocates were examined for emerging biomedical concepts and changes in domain-specific usage. Coordinated adjectives and modifying adverbs were analysed separately to investigate changes in the evaluative and promotional phraseology surrounding *novel*.

4. Results

4.1 Broadening of the collocational profile

The collocational profile of *novel* broadened over the 36-year study period. Among the 264 noun collocates occurring at least 100 times across five or more years, 110 (41.7%) *first* emerged after 1985. Furthermore, 242 (91.7%) showed increasing Collocate Preference, whereas

only 104 (39.4%) showed increasing Collocate Prominence. Together, these patterns indicate that *novel* became associated with an expanding range of concepts. As *new* collocates entered its collocational profile, many established collocates accounted for a smaller proportion of its total uses. The following sections examine the nature of these changing associations.

4.2 Increasing associations with abstract concepts

Among the noun collocates, 248 were included in Schmid's (2000) inventory of shell nouns. The analysis was restricted to the 69 shell nouns occurring at least 100 times, which accounted for 97.5% of all shell-noun tokens. Figure 1 groups the retained shell nouns into five broader discourse-semantic categories (Planning, Knowledge, Modality, Explanation, and Classification) derived from Schmid's (2000) semantic *framework* but reorganised to reflect their predominant functions in biomedical grant discourse.

The x-axis shows Kendall's τ for Collocate Preference and the y-axis Kendall's τ for Collocate Prominence. Positive values indicate increasing trends over time. Nouns in the upper-right quadrant therefore became both increasingly associated with *novel* and accounted for an increasing proportion of its collocational profile. Label size represents overall collocate growth, while colour indicates the period in which a noun *first* appeared as a collocate of *novel*.

The Planning category brings together shell nouns concerned with the planning, organisation, and conduct of *research* and is derived primarily from Schmid's volitional, eventive, and circumstantial subcategories. *Strategy*, *solution*, and *aim* all show positive trends in both preference and prominence, while *approach*, the most frequent noun collocate throughout the study period, increased in preference but remained *relatively* stable in prominence. Examples (1)–(3) illustrate the increasing tendency for different aspects of *research* planning to be characterised as *novel*.

- (1) AIDD is a *novel strategy* that offers many mechanisms to selectively deliver ...
- (2) A recent publication presents a *novel solution* that may be enhanced to gain ...
- (3) This *novel aim* tests the *hypothesis* that these cells are activated ...

The Knowledge category comprises Schmid's conceptual and creditive subcategories. The strongest increases occur for *discovery*, *knowledge*, *understanding*, and *insight*, while conceptual nouns such as *concept*, *construct*, and *paradigm* also become increasingly associated with *novel*. Several of these nouns *first* appeared as collocates after 1985. Examples (4)–(6) suggest partial semantic redundancy, since *insight*, *knowledge*, and especially *discovery* already imply a *new* contribution.

- (4) Our proposed studies will provide *novel* and significant *insight* into the pathogenesis of Pg infection and periodontitis ...
- (5) We will generate *novel knowledge* on ECM-cell interaction with respect to resident lung cell function and tissue regeneration in IPF.
- (6) This *novel discovery* greatly stimulates our curiosity to investigate whether apo A5 plays a previously unrecognised and *critical* role ...

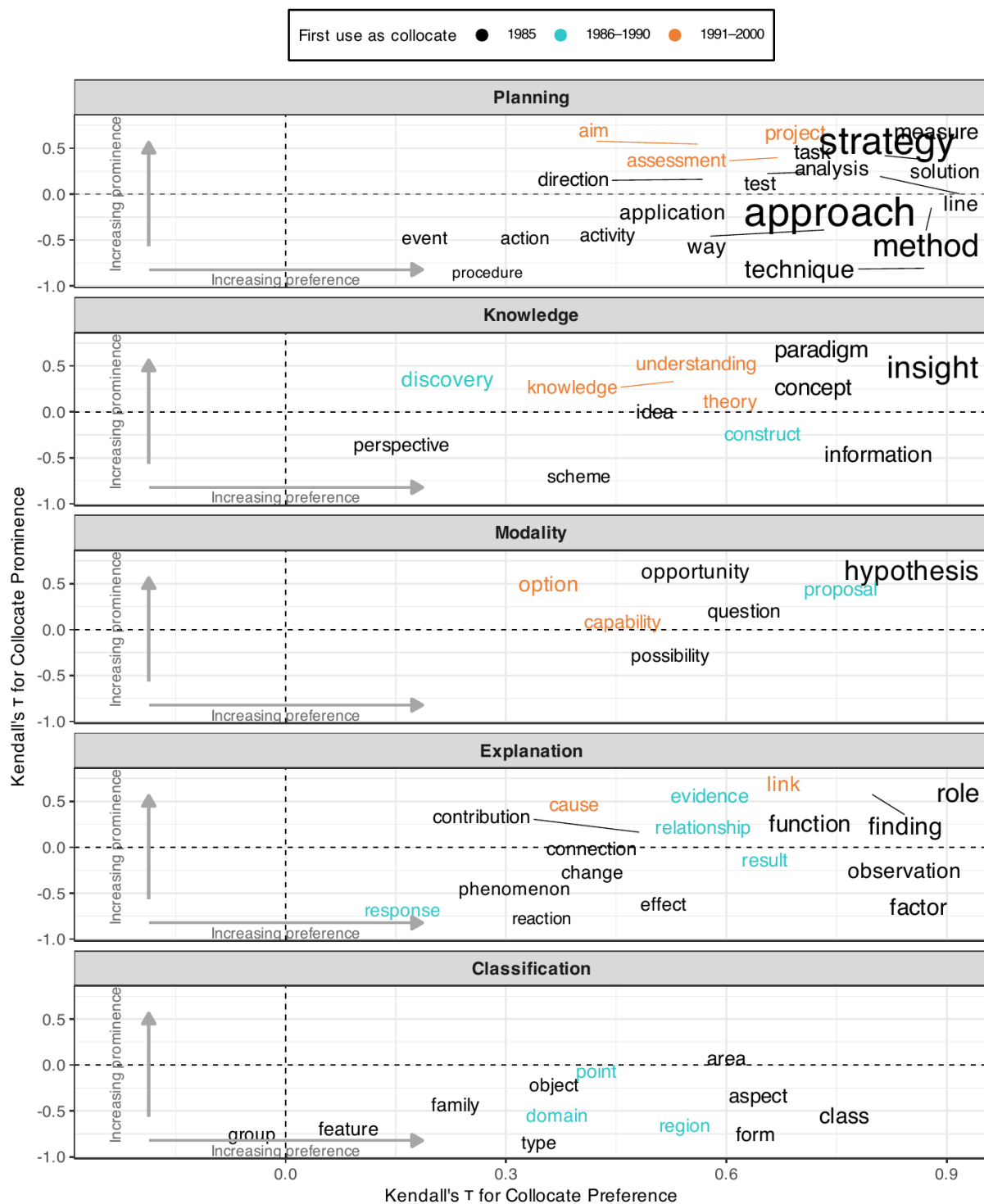


Figure 1: Diachronic trends in shell noun collocates of novel.

The Modality category is derived primarily from Schmid's modal subcategories and includes nouns expressing *possibility*, *capability*, and future action. The strongest increases occur for *capability*, *hypothesis*, *opportunity*, and *proposal*, all of which show positive trends in both preference and prominence, while *possibility* increased in preference despite declining prominence. Examples (7)–(9)

illustrate the increasing tendency for projected possibilities and anticipated contributions to be characterised as *novel*.

(7) We propose a *novel hypothesis* that by reducing inflammation, niacin treatment will restore ...

(8) The proposed *research* will lead to a *novel capability* of optically recording ...

(9) This legislation presents a *unique* and *novel opportunity* to understand ...

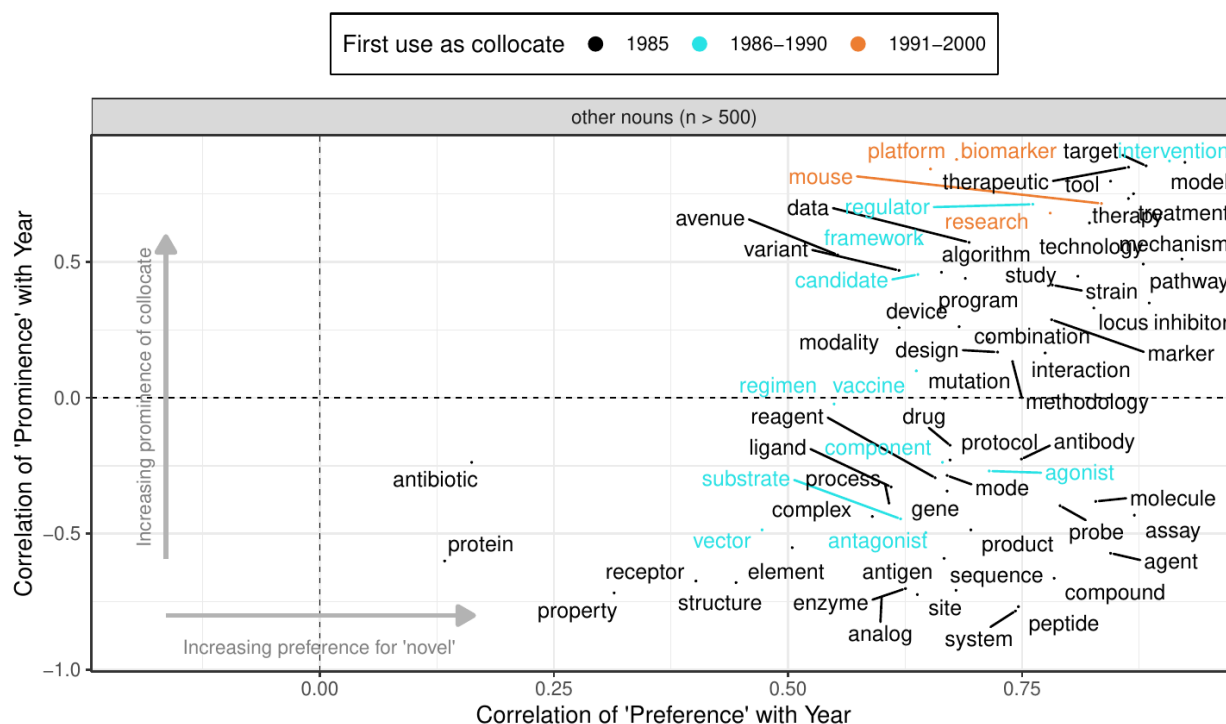


Figure 2: Diachronic trends in non-shell noun collocates of novel.

The Explanation category is derived primarily from Schmid's causal and evidential subcategories. Strong increases occur for *cause*, *evidence*, *relationship*, and *link*, all of which *first* emerged as collocates after 1985 and subsequently increased in both preference and prominence. Examples (10) and (11) illustrate the increasing tendency for explanatory and evidential relations to be characterised as *novel*.

(10) We believe that IgE to alpha-gal represents a *novel cause* of food allergy.

(11) Our finding reveals a *novel link* between autophagy and meiotic translation.

Finally, the Classification category is derived mainly from Schmid's partitive and neutral subcategories and includes nouns used to classify or delimit referents, for example by *type*, *feature*, *aspect*, or *domain*. Unlike the other categories, most show increasing preference but stable or declining prominence.

4.3 Changing associations with biomedical concepts

The non-shell collocates (Figure 2) are largely domain-specific biomedical terms (e.g. *gene*, *protein*, *receptor*, *biomarker*, and *therapeutic*). Their changing association with *novel* reflects both the emergence of new biomedical concepts and an increasing tendency for those concepts to be characterised as *novel*. The collocates appearing after 1985 correspond to successive developments in biomedical research, including molecular genetics (*gene*, *receptor*, *ligand*), technological and computational approaches (*platform*, *framework*, *algorithm*), and more recent emphases on targeted and precision medicine (*target*, *biomarker*, *therapeutic*, and *intervention*).

One interesting exception is the general noun *research*, which did not occur as a collocate of *novel* until 1993. Prior

to this, novelty was typically attributed to the products or outcomes of *research* - for example, methods, discoveries, or interventions - rather than to the *research* itself. Its later emergence therefore suggests a rhetorical shift towards presenting originality as an explicit property of the *research* as a whole.

4.4 Increasing associations with evaluative phraseology

Among the adjective collocates, 95 were classified as hype adjectives using Millar et al. (2022). We focus on the 16 occurring at least 50 times, which account for 54% (3,555/6,545) of all hype-adjective tokens. For adverbs, 14 items were retained, accounting for 56% (2,203/3,947) of all modifying adverbs. Figure 3 shows trends for the adjective and adverb collocates.

Among the adjectives, the clearest increases involve terms overlapping with *novel* in meanings of newness and innovation. *Innovative* increased in both preference and prominence, while *unique* and *first* increased in preference. Other near-synonymous adjectives, including *groundbreaking*, *revolutionary*, *unparalleled*, and *unprecedented*, also emerged as collocates during the study period. Adjectives expressing importance and value, including *key*, *critical*, *important*, and *essential*, show similar positive trends.

(1) *Novel* and *innovative* biological concepts pursued in this *proposal* include ...

(2) The specific aims are: 1) Develop a *unique novel* system for vaccines based on ...

(3) Together, these studies will provide *novel* and *critical* data on the impact of age ...

The co-occurrence of *novel* with semantically overlapping adjectives such as *innovative* and *unique* in (1) and (2) suggests reinforcement of the novelty claim, while (3)

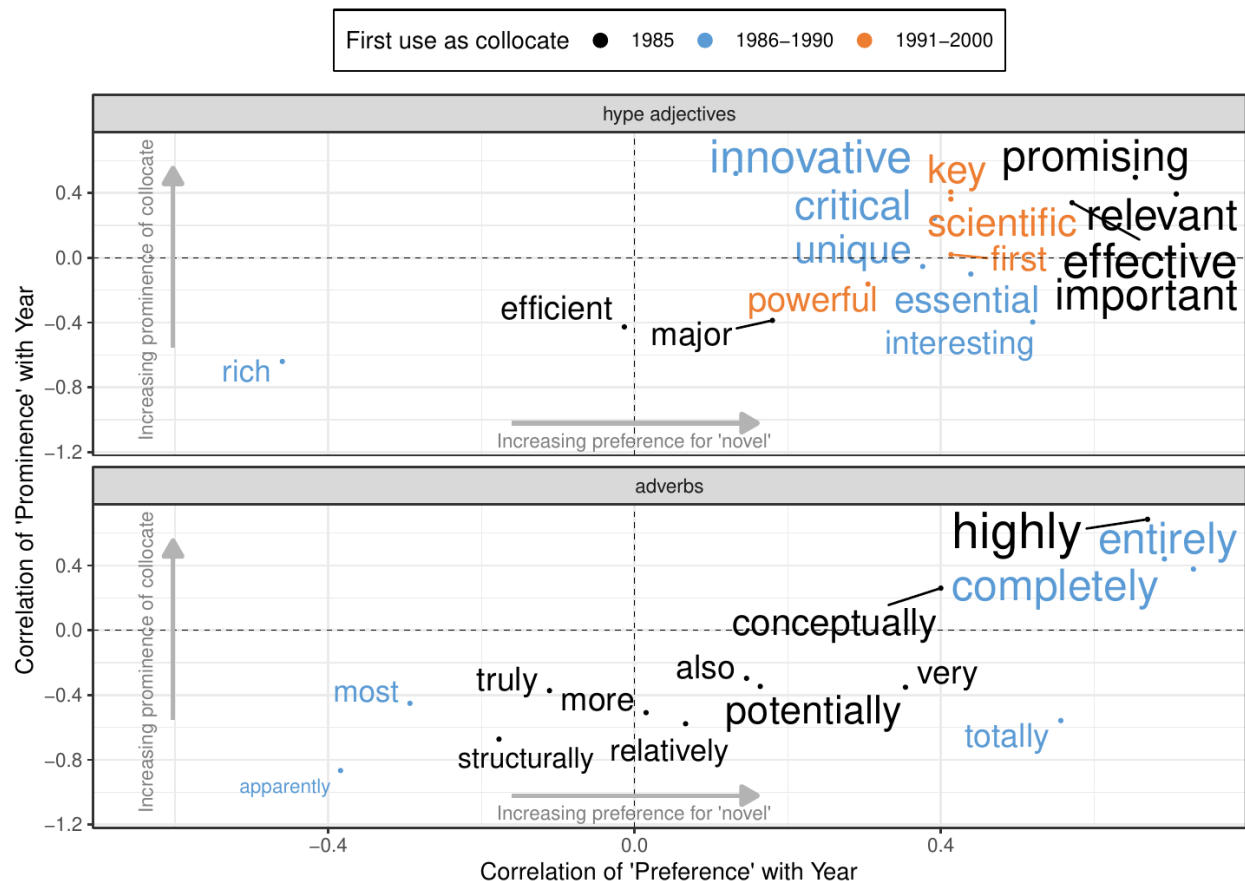


Figure 3: Diachronic trends in hype adjective and modifying adverb collocates of novel.

combines novelty with an explicit evaluation of importance. The pattern is consistent with increasingly evaluative phraseology surrounding *novel*.

The adverb collocates show a related shift towards intensification. *Highly*, *completely*, and *entirely* all show positive trends in preference and prominence, whereas weaker or qualifying modifiers such as *relatively* and *potentially* show declining prominence.

(4) To address this problem, we propose a *completely novel* removal strategy by transforming biofilms ...

(5) Ultimately, the goal is to discover an *entirely novel* therapeutic class of anti-virulence agents ...

Examples (4) and (5) treat *novel* as gradable and strengthen its force through degree modification. Together with the increasing coordination of *novel* with near-synonymous and evaluative adjectives, this pattern is consistent with rhetorical intensification and may also be compatible with some weakening of the adjective's independent semantic force.

Taken together, the above analyses indicate that the increasing use of *novel* was accompanied by systematic changes in its collocational profile. The adjective became associated with a broader range of abstract and biomedical concepts while increasingly occurring in evaluative and intensifying phraseology. These developments are consistent with semantic broadening and may also represent early evidence of semantic bleaching. In particular, the emergence of more forceful novelty terms such as *groundbreaking*, *revolutionary*, and *unprecedented*

may have reduced the rhetorical impact of *novel*, contributing to its increasing reinforcement through semantically overlapping adjectives and degree modifiers.

4.5 Interpreting diachronic collocational change

The above case study provides an opportunity to evaluate the proposed framework. Restricting the analysis to direct dependency relations produced collocates that were generally more semantically interpretable than those derived from span-based co-occurrence, although the practicality of the analysis depended strongly on the size and diversity of the collocate set. Modifying adverbs were straightforward to examine because they formed a small and semantically coherent group, whereas noun collocates, and to a lesser extent coordinated adjectives, required substantial filtering, semantic grouping, and concordance analysis before broader patterns could be identified.

The combination of Preference and Prominence was useful in distinguishing different types of change: some collocates increased in both measures, others increased in Preference while remaining stable or declining in Prominence, and previously unattested collocates could be identified as they emerged. At the same time, the filtering criteria favoured frequent and sustained associations, making the analysis better suited to identifying expansion than decline; collocates that were becoming less frequent or disappearing were less likely to be retained.

Further evaluation should compare the proposed measures with alternative approaches, including established

association measures, and examine related lexical items such as *new* and *innovative* to determine which patterns are specific to *novel* and which reflect broader changes in biomedical discourse.

5. Conclusion

This study explored a framework for investigating diachronic semantic change through dependency-based collocation analysis. Using a corpus of 901,717 NIH-funded research project abstracts spanning 1985–2020, we examined changes in the noun, adjective, and adverb collocates of *novel* using two complementary descriptive measures.

The analyses indicated a broadening of the collocational profile of *novel*. The adjective became increasingly associated with abstract concepts represented by shell nouns, emerging biomedical concepts reflecting developments in the discipline, and evaluative phraseology involving coordinated adjectives and intensifying adverbs. Together, these patterns are consistent with usage-based accounts of semantic change, indicating both semantic broadening and possible semantic bleaching as *novel* becomes increasingly reinforced by semantically overlapping and intensifying expressions.

More broadly, the study demonstrates the potential of diachronic collocation analysis for investigating lexical semantic change. The present analysis represents one way of quantifying such change, and further work is needed to compare different measures of collocational association and change. Applying these approaches to other lexical items and domains should help establish how far changing collocational behaviour can provide evidence of gradual semantic change.

6. Acknowledgements

This study was supported by grants 21K02919 and 25K00851 from the Japan Society for the Promotion of Science.

7. Bibliographical References

- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press, Cambridge.
- Bybee, J. (2015). *Language Change*. Cambridge University Press, Cambridge.
- Evert, S. (2009). Corpora and collocations. In A. Lüdeling and M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*, Vol. 2. De Gruyter Mouton, Berlin/New York, pp. 1212–1248.
- Firth, J. R. (1957). *Papers in Linguistics 1934–1951*. Oxford University Press, Oxford.
- Fraser, V. J. and Martin, J. G. (2009). Marketing data: Has the rise of impact factor led to the fall of objective language in the scientific article? *Respiratory Research*, 10:35.
- Gries, S. Th. (2024). *Frequency, Dispersion, Association, and Keyness: Revising and Tupleizing Corpus-Linguistic Measures*. John Benjamins, Amsterdam/Philadelphia.
- Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380–409.
- Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, 2(2):243–256.
- Hilpert, M. (2012). Diachronic collocation analysis: How to use it and how to deal with confounding factors. In K. Allan and J. A. Robinson (Eds.), *Current Methods in Historical Semantics*. De Gruyter Mouton, Berlin/Boston, pp. 133–160.
- Hilpert, M. and Gries, S. Th. (2009). Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4):385–401.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge, London.
- Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. Zenodo.
- Hyland, K. and Jiang, F. K. (2016). Change of attitude? A diachronic study of stance. *Written Communication*, 33(3):251–274.
- Hyland, K. and Jiang, F. K. (2021). 'Our striking results demonstrate...': Persuasion and the growth of academic hype. *Journal of Pragmatics*, 182:189–202.
- Millar, N., Batalo, B., and Budgell, B. (2022). Trends in the use of promotional language (hype) in abstracts of successful National Institutes of Health grant applications, 1985–2020. *JAMA Network Open*, 5(8):e2228676.
- Millar, N., Mathis, B., Batalo, B., and Budgell, B. (2023). Trends in the expression of epistemic stance in NIH research funding applications: 1985–2020. *Applied Linguistics*, 45(4):658–675.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 101–108.
- Schmid, H.-J. (2000). *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. De Gruyter Mouton, Berlin/New York.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell, Oxford.
- Vinkers, C. H., Tjeldink, J. K., and Otte, W. M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ*, 351:h6467.