Applied Linguistics Papers 29/3: 2025, 64–78

The pedagogical value of register-specific semantic prosody in L2 teaching

Michał Stefan RUTKOWSKI
E-mail: michalstefanrutkowski@gmail.com,

Abstract: Semantic prosody, the evaluative meaning words acquire from their collocational patterns, is a crucial but often overlooked component of second language (L2) competence. While its importance is recognised, its variation across different registers and genres remains an underexplored area in pedagogical practice. This article argues for the systematic integration of register-specific semantic prosody into L2 pedagogy. It supports this argument with an empirical, corpus-based study designed to illustrate the practical value of this approach. The study presents a comparative analysis of the near-synonymous verbs *to cause* and *to lead to*. Using the British National Corpus 2014 (BNC2014), the semantic prosodies of these verbs were examined across seven distinct genres, including academic prose, newspapers, fiction, and e-language, through a quantitative and qualitative analysis of their most frequent collocates. The findings reveal significant and systematic differences. *To cause* consistently demonstrates a strong negative semantic prosody, though its intensity varies across registers (e.g., it is more neutral in academic prose). In contrast, *to lead to* is predominantly neutral, but its evaluative profile also shifts depending on the genre. These register-driven variations underscore that the verbs are not freely interchangeable, despite their similar core meanings. Based on these findings, the article makes pedagogical recommendations. It advocates for the explicit instruction of register-specific semantic prosody for advanced L2 learners (B2 and above), who are best positioned to benefit from such nuanced instruction. It proposes a data-driven learning (DDL) methodology, empowering learners and educators to explore authentic language patterns using a suite of free, web-based corpus tools. By integrating this approach, L2 pedagogy can move beyond denotational definitions to foster a deeper, context-aware linguistic competence.

Keywords: semantic prosody, corpus linguistics, language pedagogy, second language acquisition

Introduction

The study of how meaning is created and understood in context is a cornerstone of applied linguistics. Within this domain, the concept of semantic prosody offers a lens for examining the evaluative dimensions of lexical items. This article argues that a nuanced understanding of semantic prosody, particularly its variation across different registers and genres, is a crucial and practical component that should be integrated into second language (L2) pedagogy. We aim at making the case for pedagogical value in incorporating semantic prosody into L2 teaching, which ought to help learners achieve a more native-like and contextually appropriate command of language.

The origin of semantic prosody can be traced to the late 1980s. Although he did not coin the term, Sinclair (1987) first laid the groundwork with his analysis of the phrasal verb *to set in*, noting its tendency to co-occur with words denoting unpleasant

states. The term "semantic prosody" was popularised by Louw (1993), who drew a parallel with Firth's concept of "phonological prosody", where the pronunciation of a given sound is influenced by its immediate neighbouring sounds. Louw proposed that, similarly, a word's meaning is "coloured" by its typical lexical environment. He defined semantic prosody as "a consistent aura of meaning with which a form is imbued by its collocates" (Louw 1993), citing examples like *utterly* and *bent on* which carry an unfavourable prosody. The field quickly expanded, with scholars like Bublitz (1996) identifying similar patterns in words like *cause*, *happen*, and *commit*, which become "habitually associated" with semantically consistent sets of collocates.

As the field developed, it became necessary to define the scope and nature of the phenomenon and distinguish it from related linguistic concepts. The scope of semantic prosody is not limited to single words; it can apply to larger multi-word "lexical units" (Sinclair 1998), such as *symptomatic of*. While it has been described variously as a type of meaning or a process, for pedagogical purposes, we propose that it be most usefully framed as an evaluative *feature* of a lexical item.

Before further discussion, one ought to disambiguate semantic prosody from similar terms. It differs from *collocation* (the co-occurrence of specific words) and *colligation* (co-occurrence with grammatical categories) by being evaluative in nature (Sinclair 1998). It is also distinct from *connotation*; while both are evaluative, semantic prosody is directly observable and quantifiable in corpus data through recurring collocational patterns, whereas connotation is often more schematic and less dependent on the immediate co-text (McEnery et al. 2006; Louw 2000). Finally, it must not be confused with *semantic preference*, which describes a lexical item's co-occurrence with a particular semantic set (e.g., *naked eye* with words of "visibility"), whereas semantic prosody describes the attitudinal or pragmatic function of that association (e.g., *naked eye* with a prosody of "difficulty") (Sinclair 1996; Stubbs 2001).

A central debate in the literature has concerned the "hidden" nature of semantic prosody and its accessibility to a speaker's intuition. Early researchers tended to characterize it as a covert or subliminal phenomenon, discoverable only through the computational analysis of large corpora. Stubbs (1995), for instance, argued that "native speaker intuitions are not a reliable source of evidence", a view echoed by Tognini-Bonelli (2001), who suggested that prosodies "operate mainly subliminally". However, a growing body of psycholinguistic research has challenged this view. Studies have shown that semantic prosody influences affective priming (Ellis/ Frey 2009) and affects judgments about ambiguous information (Hauser/ Schwarz 2016, 2018), demonstrating that language users are, at least implicitly, sensitive to it. Furthermore, experiments have confirmed that speakers can consciously access this knowledge. Participants in studies by Nordquist (2004) and McGee (2012) were able to produce sentences that correctly reflected the negative semantic prosody of words like *cause*. Finally, a landmark study by Stempel (2019) tested explicit knowledge directly and concluded that "speakers generally possessed better explicit knowledge of semantic prosody than assumed in the prior literature". This aligns with findings that this knowledge can be acquired through both incidental immersion and explicit training

(Guo et al. 2011), making a strong case that semantic prosody is indeed a teachable concept.

While the primary framework for describing semantic prosody has been the positive-negative polarity, some argue that it fails to capture the full complexity of evaluative meaning. Recognising the subjectivity inherent in such labels (Stewart 2010), scholars have proposed more granular descriptive models. These include references to *Appraisal Theory* (Martin/ White 2005), which deconstructs evaluation into categories of Attitude, Graduation, and Engagement; *Osgood's Evaluative Scales* (Osgood et al. 1957), which provide a quantitative basis for analysis (Dilts/ Newman 2006); and *Bednarek's (2008) Evaluative Parameters,* which offer a multi-dimensional system including concepts like comprehensibility, expectedness, and reliability. Furthermore, research has shown that some words can exhibit "dual prosody", having both positive and negative associations depending on the syntactic context, as is the case with the word *challenge* (Lin/ Chung 2016). Nevertheless, the positive-negative polarity remains the most frequently adopted framework, which allows for convenient comparison between lexis.

Finally, two critical dimensions of variation underscore the need for a nuanced, context-aware approach to teaching semantic prosody: cross-linguistic differences and register-specificity. Firstly, it cannot be assumed that near-synonyms across different languages will share the same semantic prosody. Comparative studies have revealed significant divergences; for example, Partington (1998) found that the English word *impressive* carries a generally positive prosody, while its Italian cognate *impresionante* tends towards negativity. Similarly, Ebeling (2013) demonstrated that the English verb *to cause* is strongly negative, whereas its Norwegian counterpart *forårsake* is largely neutral. Secondly, and most crucially for this article, a word's semantic prosody can vary significantly across different registers and genres within the same language. As Partington (2004) noted, "it is highly likely that the quality and strength of the prosody of a good many items will differ from genre to genre". For example, the strong negative prosody of *to cause* has been shown to be "smoothed" or neutralised in the objective context of scientific academic writing (Hunston 2007; Louw/ Chateau 2010), and the predominantly negative verb *to undergo* becomes neutral in technical and scientific English (Stubbs 2001).

In summary, semantic prosody is a multifaceted and dynamic feature of lexical meaning. It is empirically demonstrable, intuitively accessible, and describable with increasing nuance. Its variation across languages and, most importantly, across registers, presents both a challenge and an opportunity for L2 pedagogy. Learners who rely solely on denotational (dictionary) definitions are often unaware of these evaluative nuances, leading to pragmatic errors and unnatural-sounding language—a frequent issue, given that L2 learners often struggle with collocations (Namvar et al. 2012) and that textbooks often fail to represent this feature accurately (Zhang 2009; Lee 2006). Therefore, incorporating register-specific semantic prosody into L2 instruction is essential for bridging the gap between lexical knowledge and proficient, contextually sensitive language use.

In the next section, we present a study designed to empirically demonstrate these theoretical claims by analysing how the semantic prosodies of the near-synonymous verbs *to cause* and *to lead to* shift across different genres.

1. Methodology

The empirical component of this article involves a corpus-based study designed to analyse and compare the register-specific semantic prosodies of the near-synonymous verbs *to cause* and *to lead to*. The methodology was designed to be systematic and replicable, combining quantitative collocational data with qualitative analysis.

1.1 Corpus and tools

The primary data source for this study was the *British National Corpus 2014 (BNC2014)*, a 102-million-word corpus of contemporary British English. Its comprehensive nature and clear genre categorisations make it an ideal resource for investigating register-specific linguistic phenomena.

The analysis was conducted using *LancsBox X 5.0.3* (Brezina/ Platt, 2024). Concordance lines and collocational data were extracted from the BNC2014 general corpus and six distinct sub-corpora: *academic prose* (20m words), *e-language* (5m words), *fiction* (20m words), *magazines* (15m words), *newspapers* (20m words), and **o***fficial documents* (7m words).

1.2 Analytical approach: collocation and KWIC analysis

To isolate the target verbs, specific search queries were used. For *to lead to*, the query [hw="lead" pos="V.*"] to was employed to find all verbal forms of the lemma "lead" followed by the preposition "to", thus filtering for its causative meaning. For *to cause*, the query was [hw="cause" pos="V.*"], capturing all verbal forms of the lemma.

For each verb in each of the seven corpora, the *GraphColl* tool within *LancsBox* was used to identify the 30 most frequent collocates. The collocational window was set to its maximum range of 10 words to the left and 10 to the right (L10-R10) to capture a broad contextual environment. From the generated lists, irrelevant items such as articles, pronouns, prepositions, numerals, and high-frequency function verbs (e.g., forms of *to be*, *to have*) were manually excluded to focus the analysis on semantically meaningful collocates. A key methodological decision was to classify modal verbs (e.g., *can, will, may*) and words of comparison or degree (e.g., *more, increase, high*) as neutral, following frameworks that treat them as markers of possibility/necessity or graduation rather than inherently positive or negative evaluation (Bednarek 2008).

1.3 Statistical measures and qualitative evaluation

Three primary statistical measures were used for the quantitative analysis:

1. *Frequency of Occurrence:* The raw count of each collocate.
2. *Normalised Frequency (NF)*: Calculated per million words (raw frequency/ total words in corpus) * 1,000,000 to allow for objective comparison of frequencies across sub-corpora of different sizes.

3. *LogDice Score:* A statistical measure of collocational strength provided by *LancsBox*. Higher scores indicate a stronger, more statistically significant association between the node word and the collocate.

Following the quantitative extraction, a qualitative analysis was performed. Each of the top 30 relevant collocates for each verb and genre was categorised as *positive*, *negative*, or *neutral*. This classification was informed by the evaluative frameworks discussed in the literature review, including Appraisal Theory (Martin/ White, 2005), Osgood's scales (Osgood et al. 1957), and Bednarek's (2008) parameters. The aggregated percentages of positive, negative, and neutral collocates, along with their cumulative normalised frequencies and *LogDice* scores, formed the basis for determining the overall semantic prosody in each register.

2. Comparative analysis of *to lead to* vs. *to cause* across genres

The analysis revealed distinct and register-sensitive attitudinal profiles for the two verbs. While both function as causative verbs, their evaluative connotations and contextual applications differ significantly.

2.1. General Corpus

In the general BNC2014, *to cause* (NF: 160.70) is slightly more frequent than *to lead to* (NF: 145.93). Their semantic prosodies are markedly different. *to cause* exhibits a strong negative prosody, with 43.33% of its top collocates being negative (*damage, problems, pain, harm, death*). In stark contrast, *to lead to* is overwhelmingly neutral (86.66%), with a slight positive tendency (9.99% positive vs. 3.33% negative). This initial finding suggests a low degree of general synonymy.

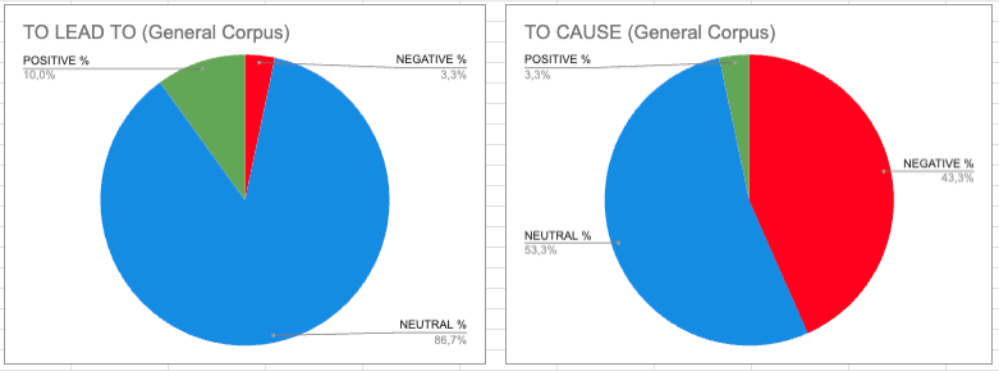

Fig. 1. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the general corpus of BNC2014.

2.2. Academic prose

The frequency pattern is reversed in academic writing, where *to lead to* (NF: 373.65) is significantly more frequent than *to cause* (NF: 247.1). The prosodies also shift. *To lead to* becomes even more neutral (90%), with a third of its collocates being words of comparison and graduation (*increase, higher, reduction*). This reflects the genre's emphasis on objective description. *To cause* also becomes more neutral here (76.66%)

than in the general corpus – a finding consistent with the "smoothing" effect often observed in objective registers – though its negative component (20%) is still six times higher than that of *to lead to* (3.33%). This suggests a higher degree of interchangeability in neutral academic contexts, but *to cause* remains the preferred choice for denoting explicitly adverse outcomes.

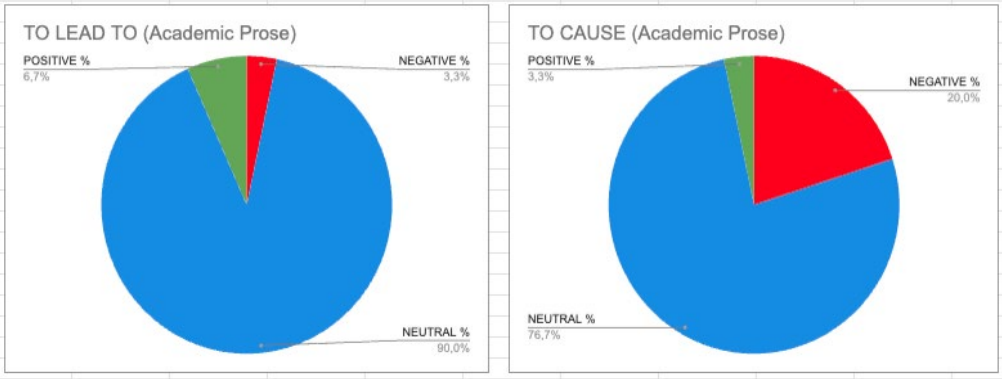

Fig. 2. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the academic prose subcorpus of BNC2014.

2.3. E-language

In the informal and unregulated context of e-language, *to cause* (NF: 158.8) is almost 2.5 times more frequent than *to lead to* (NF: 64.6). Here, the verbs show their most similar profiles. *To lead to* has an equilibrium of negative and positive collocates (13.33% each), creating ambiguity. *To cause* retains a strong negative prosody (33.33%), though less pronounced than in the general corpus. Both verbs share many neutral collocates (*can, will, more, much*). This convergence suggests a blurring of their distinct functions in informal online communication, perhaps due to the presence of non-native or less-than-proficient users in the discourse.

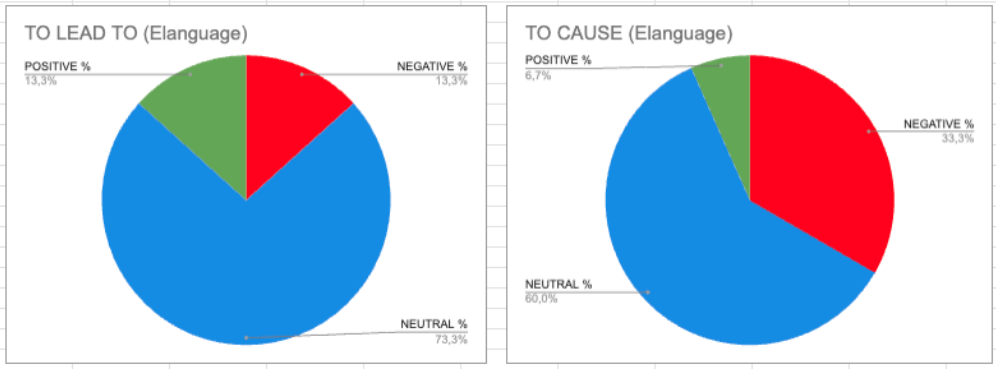

Fig. 3. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the e-language subcorpus of BNC2014.

2.4. Fiction

In fiction, *to cause* (NF: 111.55) is more than twice as frequent as *to lead to* (NF: 48.4). Their semantic profiles are highly distinct. *To lead to* is predominantly neutral (90%), but a significant portion of its collocates refer to literal, physical movement and destinations (*door, road, corridor, path*). *To cause*, conversely, lacks this literal sense and maintains a mild negative prosody (20% negative collocates like *trouble, damage, pain*), with a complete absence of positive collocates. This functional divergence makes them virtually non-synonymous in narrative contexts.

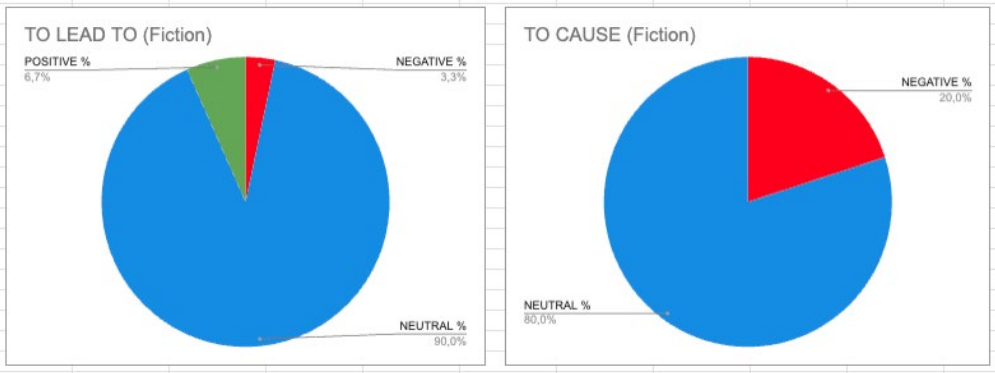

Fig. 4. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the fiction subcorpus of BNC2014.

2.5. Magazines

To cause (NF: 164.93) is more frequent than *to lead to* (NF: 134.86). *To lead to* is predominantly neutral (90%) and exhibits a clear semantic preference for collocates related to economy and finance (*prices, growth, profit, cash*). *To cause* is also largely neutral (83.33%) but has twice the proportion of negative collocates (13.33% vs. 6.66%) and a less defined semantic preference. The differing semantic profiles indicate a low degree of synonymy.

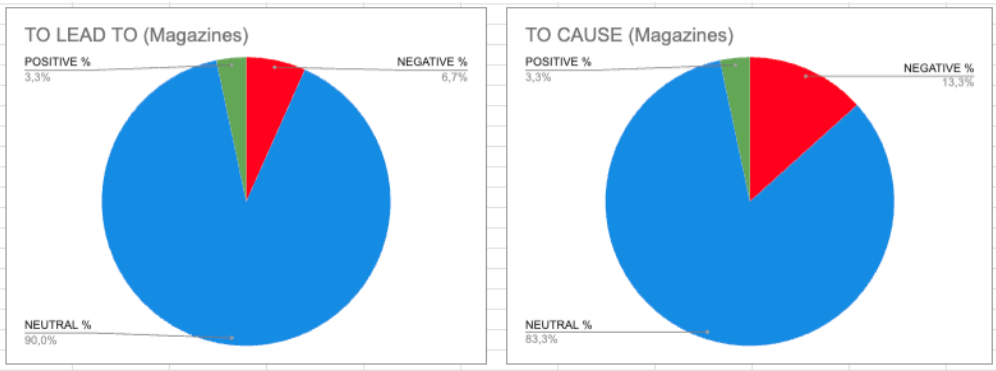

Fig. 5. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the magazines subcorpus of BNC2014.

2.6. Newspapers

To lead to (NF: 146.15) is almost three times more frequent than *to cause* (NF: 54.2), suggesting it is preferred for general reporting of consequences. Interestingly, the evaluative profile of both verbs leans towards negativity. *To lead to* is mostly neutral (80%), but its negative collocates (13.33%) outweigh the positive ones. *To cause* is much more strongly and specifically negative (26.66%), co-occurring with words denoting physical or psychological harm (*damage, harm, pain, distress, offence*), whereas *to lead to* collocates with words of potential or abstract loss (*death, loss, risk*).

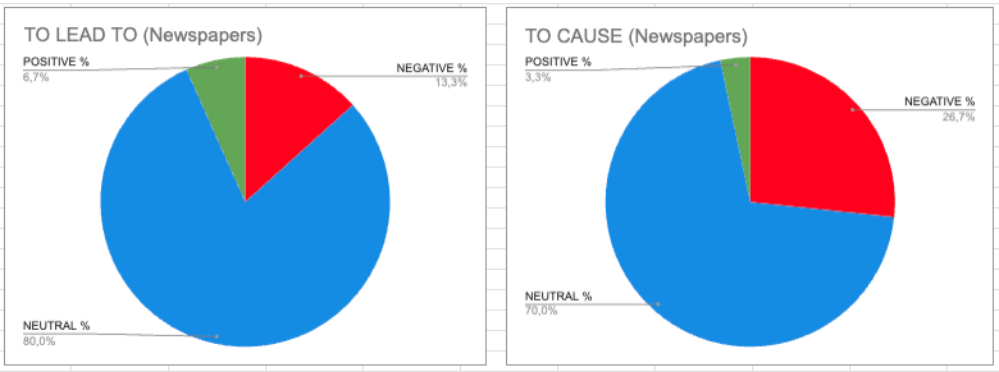

Fig. 6. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the newspapers subcorpus of BNC2014.

2.7. Official Documents

The verbs appear with similar frequency. *To lead to* is highly neutral (90%), consistent with its use in procedural and descriptive contexts. *To cause* is also majority neutral (70%) but has a powerful negative component (23.33%) and, uniquely, a notable positive component (6.66%) with high collocational strength (*significant, free*). Both verbs share a semantic preference for financial and institutional language, but the negative collocates of *to cause* specifically relate to procedural failures (*error, fraud, misstatement, risk*), clearly distinguishing its function.

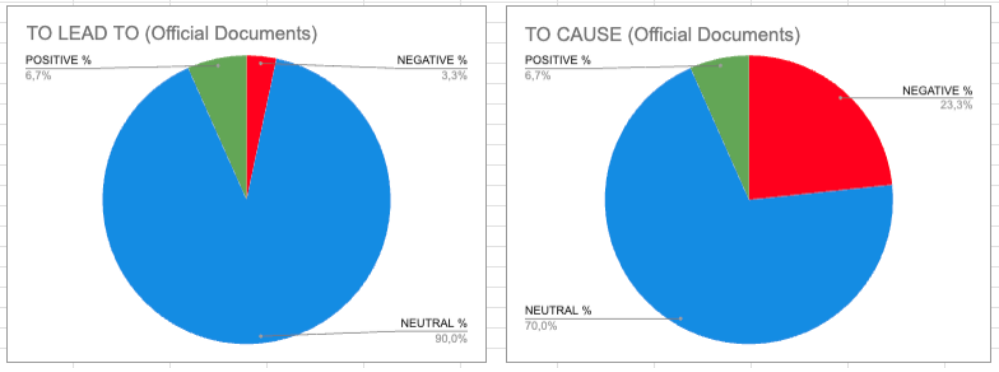

Fig. 7. The comparison of the evaluative load of the top 30 collocates of the verbs "to lead to" and "to cause" in the official documents subcorpus of BNC2014.

2.8 Summary of Findings Across Registers

The comparative analysis reveals clear and consistent differences in the semantic prosodies of *to cause* and *to lead to*, which are influenced by genre.

The verb *to cause* consistently carries a negative semantic prosody across all registers. However, the intensity of this negativity is variable. It is most pronounced in the general corpus (43.3% negative collocates) and newspapers (26.66%), where it associates with direct harm and problems. In contrast, its negativity is significantly "smoothed" or tempered in the objective context of academic prose (20%), where it functions as a more neutral descriptor of causation.

Conversely, *to lead to* is predominantly neutral across most genres, particularly in academic prose (90%), fiction (90%), and official documents (90%). Its evaluative profile, however, shows greater flexibility than *to cause*. For instance, in fiction, it often acquires a literal sense of physical direction. In newspapers, it takes on a slightly negative meaning (13.33% negative collocates), while in the informal context of e-language, its profile becomes ambiguous, with an equal number of positive and negative collocates (13.33% each).

These register-driven variations underscore that the two verbs are not interchangeable and are governed by distinct contextual conventions.

3. Discussion and Pedagogical Implications

The comparative analysis of *to cause* and *to lead to* presented in the previous section provides clear, empirical evidence for a core argument of this article: semantic prosody is a dynamic, register-sensitive feature of lexis that has profound implications for second language (L2) acquisition. The findings demonstrate that even near-synonyms with similar denotational meanings are not freely interchangeable. Their use is governed by nuanced evaluative patterns that shift significantly across communicative contexts. For instance, the negative prosody of *to cause* was significantly tempered in the objective context of academic prose (20% negative collocates) but amplified in the evaluative language of newspapers (26.66%). This finding of register-specific variation, particularly the neutralisation of *to cause* in academic writing, directly supports earlier observations by scholars such as Hunston (2007) and Louw and Chateau (2010), who noted similar “smoothing” effects in such objective contexts. This reality stands in contrast to the simplified definitions often found in dictionaries and textbooks, which frequently fail to capture these crucial pragmatic and collocational nuances (Zhang 2009; Lee 2006).

This gap between lexical theory and pedagogical practice highlights an area for improvement. For L2 learners to move beyond basic communication towards a more native-like proficiency, they must develop a sensitivity to these evaluative “hues”. This section outlines the pedagogical implications of these findings, arguing for the systematic integration of semantic prosody into L2 instruction for both educators and learners, particularly at advanced proficiency levels.

3.1 The Target Audience: Empowering Educators and Learners

The responsibility for bridging this pedagogical gap rests on two groups: educators and learners. First and foremost, L2 educators must be equipped with a solid understanding of semantic prosody themselves. As Kemp and Timms (2022) note, the concept is typically "largely subconscious for the proficient language user and rarely taught" to teachers. Without explicit training, teachers may unknowingly perpetuate the limitations of standard textbooks or be unable to explain why a student's grammatically correct sentence sounds "unnatural". By incorporating semantic prosody into teacher training programs, educators can be empowered to provide more accurate feedback, design more effective materials, and guide learners in their own linguistic explorations.

The ultimate beneficiaries, however, are L2 learners. Research consistently shows that non-native speakers are more prone to making errors in lexical collocation than in grammar (Namvar et al. 2012). Instruction in semantic prosody directly addresses this issue. This instruction is most impactful for advanced learners (B2 and above). According to the Common European Framework of Reference (CEFR), learners at pre-B2 levels are focused on acquiring foundational vocabulary and grammatical structures. Introducing an abstract concept like semantic prosody at this stage could lead to information overload and be counterproductive (Masrek/ Baharuddin 2023). However, learners at the B2 level are beginning to engage with complex texts and develop a nuanced linguistic competence. It is precisely at this stage that they need to refine their lexical choices. For C1 and C2 learners, whose goals explicitly include understanding "implicit meaning" and differentiating “finer shades of meaning”, a conscious grasp of semantic prosody is not just beneficial – it is essential for achieving their proficiency goals (Council of Europe, 2001).

3.2 The Content: A Focus on High-Frequency Vocabulary

We propose that when deciding which lexical items to focus on, one ought to follow the principle of utility. It is most effective to integrate the study of semantic prosody with high-frequency word lists. The rationale for this is grounded in well-established linguistic principles. Zipf's Law (Moreno-Sánchez et al. 2016) demonstrates that a small number of words account for a large proportion of language use. By focusing on the semantic prosodies of the most common words – such as those found in the General Service List (GSL) or Coxhead’s (2007) Academic Word List (AWL) – educators can ensure that learners are investing their time in vocabulary that offers the highest return in terms of comprehension and production. Learning the nuanced behaviour of a high-frequency word will impact a learner’s language use far more significantly than learning about a rare one.

3.3 The Method: Data-Driven Learning with Free, Web-Based Tools

The most effective way to teach semantic prosody is to move away from passive memorisation and towards active, data-driven learning (DDL), an approach that has been shown to be highly effective for this purpose (Mansoory/ Jafarpour 2014). By engaging directly with authentic language data, learners can discover collocational patterns

and evaluative meanings for themselves. Fortunately, a range of powerful and free web-based corpus tools makes this approach accessible to any classroom or independent learner.

1. Lancaster University Online Concordancer: This is an exceptionally valuable tool due to its user-friendly interface and, most importantly, its built-in filter for analysing language across different genres (e.g., *Academic, Newspapers, Fiction, Elanguage*). It allows learners to directly replicate the kind of register-specific analysis conducted in this study, making it possible to see firsthand how the prosody of a word like *to cause* shifts between a novel and a scientific article.
2. WebCorp Concordancer: Using the entire World Wide Web as its corpus, this tool provides access to a massive and current dataset in over 40 languages. Its key feature is the ability to limit searches to specific websites, allowing educators or learners to create ad-hoc analyses of particular genres (e.g., political news sites, fan forums, etc.).
3. Sketch Engine for Language Learning (SKELL): Designed specifically with learners in mind, SKELL offers a simplified interface that provides example sentences (KWIC), a "Word Sketch" (a one-page summary of a word's grammatical and collocational behaviour), and a list of similar words. It is an excellent starting point for learners new to corpus analysis.
4. NoSketchEngine and KonText: These are more advanced platforms that offer a wider array of corpora and more sophisticated search capabilities, such as filtering by part of speech (a feature used in this study). While they may require more initial guidance, they empower learners to conduct more precise and powerful linguistic investigations.

By deploying these tools, learners are no longer simply presented with a conclusion that a certain word is "negative"; they can explore the evidence for themselves in hundreds of authentic examples. This process of active discovery fosters critical thinking, enhances memory retention, and equips learners with the analytical skills to continue exploring language independently long after the lesson has ended.

4. Discussion and Pedagogical Implications

This article has demonstrated that integrating register-specific semantic prosody into L2 pedagogy is not merely an enhancement but a crucial step towards fostering a deeper, more authentic linguistic competence. By tracing the concept from its theoretical origins to its practical application, we have made a case for a more nuanced approach to vocabulary instruction. This approach moves beyond denotational meaning to embrace the evaluative and pragmatic forces that shape language in use.

We began by grounding our discussion in a unified theoretical framework, drawing upon decades of research from the foundational work of Sinclair (1987) and Louw (1993) to contemporary debates. This article has situated semantic prosody as an empirically verifiable feature of lexis, distinct from related concepts like collocation and connotation. While once considered a covert phenomenon accessible only to linguists, it was argued that proficient language users possess both implicit and explicit

knowledge of it, and that this knowledge is demonstrably teachable (Stempel 2019; Guo et al. 2011). Furthermore, our review highlighted the shortcomings of a simple positive-negative polarity and underscored the importance of two key dimensions of variation: cross-linguistic differences and, most critically, register-specificity.

The empirical study of *to cause* and *to lead to* provided concrete evidence for these claims. The analysis revealed that these near-synonyms are far from interchangeable. Their semantic prosodies diverge significantly, with *to cause* maintaining a consistently stronger negative load, while *to lead to* leans towards neutrality. Crucially, these evaluative tendencies were shown to be fluid, shifting in response to the conventions of different genres. The negative prosody of *to cause* was tempered in the objective environment of academic prose but amplified in the evaluative context of newspapers, demonstrating precisely the kind of nuanced, context-dependent behaviour that L2 learners are to master.

These findings lead to clear pedagogical recommendations. We advocate for the explicit instruction of semantic prosody for advanced learners (B2 and above), who are cognitively ready to engage with these abstract concepts and whose proficiency goals demand such a nuanced understanding. By focusing on high-frequency vocabulary, educators can ensure the greatest practical impact. This instruction should be facilitated through a data-driven learning (DDL) approach, empowering learners to become active investigators of language by using free, accessible web-based corpus tools like the Lancaster University Online Concordancer, SKELL, and others.

4.1 Future Directions: The Role of Artificial Intelligence

While this study provides a clear framework, it also opens up several avenues for future research. Further cross-linguistic studies and analyses of underrepresented languages and genres would continue to enrich our understanding. However, the most transformative potential for the future of this field lies in the integration of Artificial Intelligence and Large Language Models (LLMs).

A major challenge in applying corpus linguistics to pedagogy has been the complexity of the tools and the steep learning curve required for non-linguists to use them effectively (Tribble 2000). The deployment of AI-powered tools, such as *Gemini* (Google) or *ChatGPT* (OpenAI), can revolutionise this landscape. These technologies can eliminate the need for specialised software and complex query syntax, offering a user-friendly, conversational interface. Future research should explore the efficacy of these tools in several key areas:

1. Democratizing Corpus Analysis: Learners and educators could simply ask an LLM in natural language: “Show me 20 examples of how the verb *to cause* is used in medical journals. Does it usually have a negative meaning?” This would instantly provide the kind of targeted, register-specific data that currently requires significant technical skill to extract.
2. Personalised and Instantaneous Feedback: AI could function as a personal language tutor. A learner could input a sentence, and the AI could analyse it against vast corpus-derived patterns, offering immediate feedback not just on grammar, but on collocational appropriacy and semantic prosody.

3. On-Demand Corpus Creation: The limitation of being confined to pre-existing corpora could be overcome. Future AI systems could be tasked with crawling the web to create instant, specialised corpora on demand – for example, a corpus of recent political speeches or product reviews – allowing for timely and highly specific linguistic analysis.

Controlled classroom studies comparing traditional DDL approaches with these AI-driven methods would be invaluable in assessing their effectiveness. By harnessing the power of AI, we can make the insights of corpus linguistics more intuitive, accessible, and personalised than ever before. Ultimately, these AI-driven methods represent a powerful means of scaling the data-driven learning approach advocated in this paper, making the nuanced, context-aware instruction of semantic prosody accessible to all learners.

By bridging the gap between linguistic theory and pedagogical practice, and by looking ahead to the transformative potential of new technologies, this research lays the groundwork for a more refined and authentic approach to language education in the 21st century.

Bibliography

Bednarek, M. (2008), *An increasingly familiar tragedy*. Evaluative collocation and conflation, (in:) "Functions of Language" 15/1, 133–158.

Brezina, V./ W. Platt (2024), *#LancsBox X* [software], Lancaster University. URL: http://lancsbox.lancs.ac.uk. [Accessed 1.10.2024].

Bublitz, W. (1996), *Semantic prosody and cohesive company: somewhat predictable*, (in:) "Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie".

Council of Europe. (2001), *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge.

Dilts, P./ J. Newman (2006), *A note on quantifying "good" and "bad" prosodies,* (in:) *"*Corpus Linguistics and Linguistic Theory" 2(2), 233–242.

Ebeling, S.O. (2013), *Semantic prosody in a cross-linguistic perspective,* (in:) M. Huber/ J. Mukherjee (eds.), Corpus Linguistics and Variation in English: Focus on Non-Native Englishes, vol. 13, online. Studies in Variation, Contacts and Change in English.

Ellis, N.C./ E. Frey (2009), *The psycholinguistic reality of collocation and semantic prosody (2): Affective priming,* (in:) R. Corrigan (eds.), Formulaic language, Volume: Acquisition, loss, psychological reality, and functional explanations, volume 83 of Typological studies in language, John Benjamins.

Guo, X. et al. (2011), *Acquisition of conscious and unconscious knowledge of semantic prosody,* (in:). "Consciousness and Cognition" 20(2), 417–425.

Hauser, D.J./ N. Schwarz (2016), *Semantic prosody and judgment*, (in:) "Journal of Experimental Psychology: General" 145(7), 882–896.

Hauser, D.J./ N. Schwarz (2018), *How seemingly innocuous words can bias judgment: Semantic prosody and impression formation,* (in:) *"*Journal of Experimental Social Psychology" 75, 11–18.

Hunston, S. (2007), *Semantic prosody revisited*, (in:) "International Journal of Corpus Linguistics" 12(2), 249–268.

Kemp J./ L. Timms (2022), *Exploring semantic prosody with trainee teachers. Teaching English with Corpora: A Resource Book.* Routledge.

Lee, J. (2006), *Subjunctive were and indicative was: A corpus analysis for English language teachers and textbook writers,* (in:) "Language Teaching Research" 10(1), 80–93.

Lin, Y./ S. Chung (2016), *A corpus-based study on the semantic prosody of challenge,* (in:) "Taiwan Journal of TESOL" 13(2), 99–146.

Louw, B. (1993), *Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies*, (in:) M. Baker/ G. Francis/ E. Tognini-Bonelli (eds.), Text and Technology: In Honour of John Sinclair. Amsterdam and Philadelphia.

Louw, B. (2000), *Contextual prosodic theory: bringing semantic prosodies to life*, (in:) C. Heffer/ H. Saunston (eds.), Words in Context: In Honour of John Sinclair. Birmingham.

Louw, B./ C. Chateau (2010), *Semantic prosody for the 21st century: Are prosodies smoothed in academic context? A contextual prosodic theoretical perspective*, (in:) "Statistical Analysis of Textual Data: Proceedings of the tenth JADT Conference".

Mansoory, N./ M. Jafarpour (2014), *Teaching Semantic Prosody of English Verbs Through the DDL Approach and its Effect on Learners' Vocabulary Choice Appropriateness in a Persian EFL Context,* (in:) "Advances in Language and Literary Studies" 5(2), 149–161.

Martin, J.R./ P.R.R. White (2005), *The Language of Evaluation.* New York.

Masrek. M.N./ M.F. Baharuddin (2023), *Screens, Streams, and Stress: A Qualitative Study on How Distance Learning Students Cope with Information Overload,* (in:) *"*International journal of membrane science and technology" 10(5), 47–58.

McEnery, A./ R. Xiao/ Y. Tono (2006), *Corpus-based Language Studies: An Advanced Resource Book.* London and New York.

McGee, I. (2012), *Should we teach semantic prosody awareness?,* (in:) "RELC Journal" 43(2), 169–186.

Moreno-Sánchez, I./ F. Font-Clos/ A. Corral (2016), *Large-scale analysis of Zipf's Law in English texts,* (in:) "PLOS ONE" 11(1), e0147073.

Namvar, F./ N. Mohd Nor/ N. Ibrahim/ J. Mustafa (2012), *Analysis of collocations in the Iranian postgraduate students' writings,* (in:) "3L: Southeast Asian Journal of English Language Studies" 18(1).

Nordquist, D. (2004), *Comparing elicited data and corpora,* (in:) M. Achard/ S. Kemmer (ed.), Language, culture and mind.

Osgood, Ch.E./ G.H. Suci/ H. Tannebaum (1957), *The Measurement of Meaning.* Urbana.

Partington, A. (1998), *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam.

Partington, A. (2004), *Utterly content in each other's company: semantic prosody and semantic preference*, (in:) "International Journal of Corpus Linguistics" 9/1, 131–156.

Sinclair, J.M. (1987), *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. HarperCollins Publishers Limited.

Sinclair, J. (1996), *The search for units of meaning*. Textus 9.

Sinclair, J. (1998), *The lexical item,* (in:) E. Weigand (ed.), Contrastive Lexical Semantics. Amsterdam.

Stempel, P. (2019), *A Constructional Reanalysis of Semantic Prosody.* [Doctoral thesis, Rice University].

Stewart, D. (2010), *Semantic Prosody: A Critical Evaluation*. Routledge Advances in Corpus Linguistics.

Stubbs, M. (1995), *Collocations and semantic profiles: On the cause of the trouble with quantitative studies,* (in:) "Functions of Language" 2(1), 23–55.

Stubbs, M. (2001), *Words and phrases: Corpus studies of lexical semantics.* Oxford.

Tognini-Bonelli, E. (2001), *Corpus Linguistics at Work.* Amsterdam.

Tribble, C. (2000), *Genres, keywords, teaching: towards a pedagogic account of the language of project proposals,* (in:) L. Burnard/ A. McEnery (eds.), Rethinking Language Pedagogy from a Corpus Perspective. Papers from the third international conference on Teaching and Language Corpora. Frankfurt.

Zhang, W. (2009), *Semantic prosody and ESL/EFL vocabulary pedagogy,* (in:) "TESL Canada Journal" 26(2), 1–12.