



Research on French Semantic Prosody Based on the Model of Extended Units of Meaning

YANG Tong^{1*}, CAO Yuyue²

¹²School of Foreign Languages, North China Electric Power University, Beijing 102206, China

*Corresponding Author Email: tongyang@ncepu.edu.cn

DOI: [10.33329/rjelal.13.4.168](https://doi.org/10.33329/rjelal.13.4.168)



Article info

Article Received: 03/10/2025
Article Accepted: 07/11/2025
Published online: 13/11/2025

Abstract

Semantic prosody is one of the most important concepts in corpus linguistics, and semantic prosody research has opened up a new direction for corpus linguistics research and played an important leading role in language research. This paper takes the model of extended units of meaning as the theoretical framework to study French semantic rhyme and explores the general method of French vocabulary research through case analysis. Research shows that the model of extended units of meaning provide a sophisticated and powerful analytical tool for French vocabulary research, advancing our understanding of French vocabulary from the isolated "lexical meaning" level to the overall "meaning-function" level.

Key words: semantic prosody; extended units of meaning; Collocation; Class conjunction; French vocabulary.

1.Introduction

J.R. Firth, a renowned British linguist and the founding figure of British linguistics, established the London School of Speech. His theories, known as Firthian Linguistics among Western scholars, have had a profound influence on the field. The concept of semantic prosody, a significant discovery in corpus linguistics, can trace its origins back to Firthian Linguistics. Influenced by the prosodic analysis developed by the phonetician Henry Sweet and the "meaning in context" proposed by social anthropologist Bronisław Malinowski, Firth introduced the contextual theory of meaning, also referred to as the theory of context of

meaning, as well as prosodic phonology. John Sinclair, the founder of corpus linguistics, further developed Firth's ideas and was the first to provide a clear description of semantic prosody.

According to Sinclair, semantic prosody "expresses attitude and lies at the pragmatic end of the semantics-pragmatics continuum," and it "plays a dominant role in integrating a word with its surroundings... what is expressed approximates to the function of an item". He also pointed out that semantic prosody represents the boundary of a lexical item. Semantic prosody refers to a pervasive attitudinal aura created by the habitual

collocations of a lexical item, typically positive, negative, or neutral. This meaning is implicit and subtle, yet it profoundly influences language comprehension and production. Currently, research on semantic prosody in English has been quite in-depth; however, in the field of French studies, this paradigm has not yet achieved the same level of popularity and application. Therefore, this study will apply the extended unit of meaning model to the investigation of semantic prosody in French, aiming to explore how to empirically examine the semantic prosody of French vocabulary using corpus linguistics methods.

2. Overview of the Extended Units of Meaning

Building upon the contextual theory of meaning, Firth further proposed five modes of meaning for linguistic study: the phonetic mode, the phonological mode, the lexical mode, the grammatical mode, and the context of culture. At the lexical level of meaning, one important mechanism for realizing meaning is through collocation. He famously asserted, "You shall know a word by the company it keeps." He believed that the meaning of a word is not only determined by the concept it refers to but also includes its habitual co-occurrence with other words in authentic contexts, known as customary collocation.

At the grammatical level of meaning, Firth introduced the important concept of colligation, which refers to "the interrelations of grammatical categories in syntactic structure". For Firth, colligation involves the co-occurrence of various grammatical choices. For example, he summarized the colligation of "I watched him" as the co-occurrence relationship of grammatical categories such as "first-person pronoun + past tense verb + third-person pronoun."

Collocation and colligation are two similar yet distinct concepts. In Firth's view, "collocation stares at you in the face just as it is," whereas colligation accumulates around a word, often exhibiting discontinuity in a

sentence, more akin to grammatical concord or agreement.

Influenced by Firthian linguistics, Sinclair proposed the concept of the extended units of meaning and established the corresponding model. This model encompasses four levels of co-selection and consists of five components: the core, collocation, colligation, semantic preference, and semantic prosody. Among these, the core serves as the center of the units of meaning, with all extensions of meaning radiating outward from it. Here, it is necessary to distinguish between the core and the node word. The node word refers to a researcher-selected word that serves as the focal point of analysis. According to Sinclair's definition, the core cannot be entirely predetermined and is not equivalent to the node word. Only after identifying the extended units of meaning and its various manifestations can the invariant components of that units be discussed and determined.

Collocation refers to the lexical patterns formed around the core, reflecting the co-selection of words with other words. Colligation denotes the grammatical structure formed by the core and its surrounding elements, representing the syntagmatic relations of the lexical item at the grammatical level, particularly in syntax. Semantic preference refers to the semantic features of the collocates, reflecting the habitual co-occurrence of the core with surrounding lexical items and summarizing the semantic characteristics of these co-occurring words. Semantic prosody reflects the semantic nature of the core itself and expresses communicative attitude.

Both semantic preference and semantic prosody reflect the semantic relationship between the core and its collocates, illustrating the pragmatic features of the core and its co-occurring items. Therefore, the five components of the extended units of meaning together achieve co-selection at the lexical, syntactic, semantic, and pragmatic levels.

3. Research Design

3.1 Research Questions

This study aims to answer three questions: First, what are the characteristics of the collocations and colligations of the word *mesure*? Second, what semantic preferences and semantic prosodies does *mesure* exhibit across different collocations and colligations? Finally, what insights can the semantic prosody analysis of *mesure* offer for French vocabulary teaching?

3.2 Research Methodology

Generally, there are three methods for extracting semantic prosody: Establishing and referencing colligations to extract semantic prosody using a data-based approach; Calculating the collocates of the node word to extract semantic prosody using a data-driven approach; Investigating semantic prosody and establishing its structure using a combination of data-based and data-driven methods. This paper primarily adopts the third method, combining data-based and data-driven approaches to study the semantic prosody of the French word *mesure*.

3.3 Research Procedure

This study should generally follow these steps: First, use the AntConc tool to search for the word *mesure* in the self-built corpus, setting a span of -4/+4 to examine the collocates within this range. Second, examine the specific contextual information of *mesure* to analyze its colligational characteristics. Given the large number of concordance lines, 150 lines were randomly selected for manual examination to establish colligations. Third, analyze the characteristics of the semantic preferences of the collocates of *mesure*. Finally, examine the 150 randomly selected concordance lines to quantify the semantic prosody information.

(1) Corpus and Node Word

This study utilizes a self-built corpus with texts selected from Frantext (French Text Corpus), Scientext (a French-led project for

analyzing scientific discourse), and Le Monde (one of France's most influential daily newspapers). The corpus size is approximately 3.62 million words, covering various domains such as literary works, academic papers, and news reports.

The node word for this study is *mesure*. The French word *mesure* has multiple meanings, making it a subject of research interest. The definitions of *mesure* from the Ninth Edition of the Dictionnaire de l'Académie française are as follows:

- 1) Détermination d'une quantité ou d'une grandeur par son rapport à une échelle de référence. (Determining a quantity or magnitude by its relationship to a reference standard.)
- 2) Grandeur conventionnelle prise comme référence pour évaluer d'autres grandeurs de même nature. (A conventional magnitude used as a reference for evaluating other magnitudes of the same nature.)
- 3) Modération, retenue qui s'accorde avec le sentiment des bienséances. (Moderation, restraint that aligns with a sense of propriety.)
- 4) Disposition prise pour arriver à un but. S'emploie surtout en parlant des décisions et des actes d'une autorité politique, administrative, judiciaire, etc. (A disposition taken to achieve a goal. Used especially when referring to the decisions and actions of a political, administrative, judicial authority, etc.)
- 5) Division de la durée d'une phrase, d'une pièce, etc. en sections égales qui sont marquées de façon plus ou moins sensible dans l'exécution. (Generally used in the field of music: The division of the duration of a phrase, a piece, etc., into equal sections, marked more or less perceptibly during performance.)

A total of 1329 concordance lines containing the French noun *mesure* were retrieved from the corpus. (Since the third-person singular present tense conjugation of the verb *mesurer* also appears as *mesure*, we

distinguished and selected only its noun form.) Then, 50 valid concordance lines were randomly selected from each of the three sub-corpora – literature, scientific texts, and newspapers – to form a mini-corpus comprising 150 concordance lines.

4. Analysis of the Extended Units of Meaning for the word mesure

(1) Register Distribution

Table 1 Distribution of the Node Word mesure Across Registers

FilePath	FileTokens	Freq	NormFreq	Dispersion
LeMonde.txt	379300	80	210.915	0.871
Scientext 1.36.txt	1366029	833	609.797	0.825
Frantext1.87.txt	1870977	416	222.344	0.790

According to Table 1 below, we can observe that the word is primarily distributed in the domain of scientific and technical articles, accounting for approximately 62% of its occurrences. This shows a certain correlation with its first two dictionary definitions. Its lower frequency in media and newspapers suggests that its usage contexts are relatively more academic and specialized.

(2) Collocate Analysis

This study utilized the AntConc tool, using mesure as the search word with a span of -4/+4,

to list the top 15 salient collocates based on frequency, as shown in Table 2 below.

Table 2 Collocates of the Node Word mesure

Collocate	Rank	Freq (Scaled)	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
fur	1	570	55	55	0	3	508.844	8.037
la	2	1073060	801	503	298	3	335.028	1.022
à	3	631340	503	416	87	3	242.233	1.116
dans	4	341770	320	253	67	3	212.733	1.349
quelle	5	6940	48	48	0	3	191.308	4.234
où	6	61180	109	6	103	3	171.780	2.277
de	7	1792210	961	424	537	3	128.755	0.545
certaine	8	2970	26	26	0	2	115.308	4.574
que	9	384960	280	70	210	3	106.761	0.985
moindre	10	2460	23	23	0	2	104.917	4.669
divise	11	390	14	1	13	1	101.087	6.610
instantanée	12	740	15	6	9	1	91.161	5.785
réparation	13	7610	29	5	24	1	83.389	3.374
erreur	14	8750	30	20	10	1	80.558	3.222
scores	15	510	12	0	12	1	76.484	

Through part-of-speech analysis of the collocates of mesure, we can categorize them into nominal collocates (fur, scores), prepositional collocates (à, dans, de), adjectival collocates (certaine, moindre), conjunctive

collocates (que, où), as well as a limited number of article collocates (la), verbal collocates (indiquent), and adverbial collocates (significativement). This demonstrates that mesure exhibits a relatively rich set of

collocations, with prepositional collocates being particularly frequent. Additionally, the high collocation frequency between *fur* and *mesure* indirectly reflects the fixed nature of the phrase *au fur et à mesure*.

(3) Observation and Identification of Colligations

By examining these 150 concordance lines, we identified a variety of colligations for *mesure*, which can be broadly classified into three types. The first type of colligation is "Preposition + (Article) + *mesure* + Conjunction", such as "*à mesure que*", "*dans la/une mesure où*", etc. For example, *au contraire, ces vides se combent à mesure que ce système se nivelle*. (On the contrary, these gaps are filled as the system stabilizes.) The second type of colligation is "Preposition + (Article) + *mesure* + Preposition", such as "*en mesure de*", "*à la mesure de*". For example, *c'est-à-dire que les apprenants soient en mesure de la comprendre*. (That is to say, learners should be capable of understanding it.) And the last type of colligation is "Verb + Article + *mesure* + Preposition", such as "*prendre la mesure de*". For example, *le premier ministre a beaucoup tardé à prendre la mesure de l'importance et de la popularité du mouvement*. (The Prime Minister was slow to grasp the scale of the movement's importance and popularity.)

Among these, the first type of colligation is the most frequent, with *à mesure que* appearing 59 times in the concordance lines, accounting for approximately 39% of the total. Although the last type of colligation occurs less frequently, its structure is relatively fixed, predominantly appearing as *prendre la mesure de*. It is worth noting that while the phrase *au fur et à mesure* is common, it is categorized here under the first two types of colligation rather than being analyzed as a separate classification.

(4) Summarizing Semantic Preferences of Collocates with Reference to Colligations

When analyzing concordance lines, we observe that the word *mesure* carries different meanings and exhibits distinct semantic preferences when collocating with different prepositions. Therefore, we can further refine the colligation categories and select several representative collocational structures for detailed analysis.

Colligation Type 1 : Preposition + (Article) + *mesure* + Conjunction. Through examination of concordance lines containing the collocational structure "*à mesure que*", we identify two primary semantic preferences: The first category describes developmental progress, commonly collocating with words such as *société* (society), *division* (division of labor), and *production* (production), as illustrated in these examples: *nous amusant toujours plus, à mesure que nos moyens d'exécution se perfectionnaient* (we amused ourselves more and more as our means of execution improved) and *à mesure que les fonctions industrielles se spécialisent davantage* (as industrial functions become more specialized). The second category indicates temporal progression, exemplified by: *Ma conviction ne devint que plus certaine, à mesure que les jours s'écoulaient*. (My conviction only became more certain as the days passed.) The collocational structure "*dans la/une mesure où*" demonstrates a strong tendency to express scope, proportion, and limitation. It frequently co-occurs with gradable adjectives such as *large* (broad/extensive), *certaine* (certain/to some extent), and *faible* (limited/slight), serving to precisely delineate the degree of applicability of a statement.

Colligation Type 2: Preposition + (Article)+ *mesure* + Preposition. The construction "*en mesure de*" exhibits a clear preference for expressing capability, possibility, and feasibility. The infinitive verbs that follow encompass a range of competencies, from concrete actions such as *comprendre*

(understand) and reproduire (reproduce) to abstract cognitive processes like comparer (compare) and proposer (propose). The construction “à la mesure de” tends to express commensurability, suitability, or matching. It frequently co-occurs with abstract concepts such as beauté (beauty), identité (identity), or complexité (complexity).

Colligation Type 3: Verb + Article + mesure + Preposition. For Colligation Type 3, our analysis focuses primarily on the construction “prendre la mesure de”. Observation of its typical right-hand collocates reveals that this colligation carries a strong cognitive and evaluative function. Notably, its object of evaluation is almost invariably a severe, negative, or highly significant abstract situation, such as catastrophe (catastrophe), désaffection (disaffection), épidémie (epidemic), or importance (importance).

(5) Inducing Semantic Prosody

Based on the semantic preferences summarized in Step 4, we can induce the semantic prosodies for the various colligation types.

Colligation Type 1 exhibits different semantic prosodies across its constituent structures: The structure “à mesure que” describes a change occurring synchronously with another process, expressing a gradual progression. Our analysis of the concordance lines reveals that this change often signifies development and progress, thus presenting a positive semantic prosody. For example: Il s'élargit de plus en plus à mesure que le travail se divise davantage.

The structure “dans la/une mesure où” serves to limit the scope or degree of a statement, typically constituting an objective description without marked emotional bias, thus demonstrating a predominantly neutral semantic prosody. For example: d'atteintes aux droits de l'homme que dans la mesure où Moscou et Pékin consentent à les dénoncer.

Analysis indicates that within the “à mesure que” structure, nearly 70% of concordance lines exhibit positive semantic prosody. Therefore, its positive semantic prosody is dominant.

Colligation Type 2, describing capability, possibility, or suitability, overall presents a positive semantic prosody. For example: c'est-à-dire que les apprenants soient en mesure de la comprendre. However, in specific contexts, a tendency towards negative semantic prosody can occur. For example: il n'a ni une gestion ni un commandement à la mesure de l'extrême complexité de fonctionnement.

Colligation Type 3, frequently co-occurring with negative events such as disasters and discontent, demonstrates a strong negative semantic prosody. For example: depuis dix jours d'avoir tardé à prendre la mesure de l'épidémie.

Furthermore, observing the semantic prosodies of these colligations across different registers reveals that Colligation Type 1 appears frequently in scientific texts with a high proportion of positive semantic prosody, whereas Colligation Type 3 is found primarily in journalistic writing and exhibits exclusively negative semantic prosody.

In conclusion, we determine that the word mesure overall presents a complex semantic prosody profile, characterized by a dominant positive prosody coexisting with both neutral and negative semantic prosodies.

5. Discussion and Conclusion

Through the analysis of 150 concordance lines, we can conclude that, overall, the word mesure does not carry a single, monolithic semantic prosody. Instead, it presents a complex spectrum dominated predominantly by a positive semantic prosody, from which negative prosodies emerge within specific grammatical structures. This demonstrates that semantic prosody is not an inherent property of a word, but rather a stable

attitudinal atmosphere that the word acquires through its habitual contextual usage.

Furthermore, the semantic prosody of this word is strongly correlated with its colligations. Within the 150 concordance lines, the colligational structure “prendre la mesure de” exhibits a negative semantic prosody in 100% of its occurrences, whereas “à mesure que” tends to present a positive prosody. The semantic prosody of “à la mesure de” can be either positive (when signifying commensurability with beauty or identity) or negative (when signifying commensurability with problems or complexity), presenting a mixed prosodic profile.

It can thus be deduced that semantic prosody is not expressed directly by a single word in isolation, but is generated through the co-occurrence and interaction between words. Therefore, the analysis of semantic prosody must incorporate contextual examination to yield accurate judgments. Mastery of semantic prosody represents an advanced level of pragmatic competence.

This study, based on the extended units of meaning model and through the analysis of the semantic prosody of the French word *mesure*, aims to explore general methodologies for researching semantic prosody in French. The results indicate that the research methods applied to semantic prosody in English lexicon are equally applicable to the study of French vocabulary.

Concurrently, this study offers a novel, more profound, and efficient perspective for French vocabulary pedagogy. It holds significant implications for improving and deepening the teaching of French vocabulary, advocating a shift from a 'vocabulary list' model towards a 'chunking' and 'pragmatic' approach. This research reveals that rote memorization of words and their Chinese translations is far from sufficient, as the same word can exhibit vastly different emotional connotations depending on its structural and contextual environment.

Consequently, we advocate using the extended units of meaning as the fundamental teaching unit and promoting the 'lexical chunk approach'. This involves learning words while simultaneously understanding and mastering their collocations, colligations, semantic preferences, and semantic prosodies. The goal is to cultivate proficient learners who can not only speak correct French but also speak authentic French.

References

- Firth, J. R. M. (1957a). *Papers in linguistics*. Oxford University Press.
- Firth, J. R. C. (1957b). Ethnographic analysis and language with reference to Malinowski's views. In R. Firth (Ed.), *Man and Culture: An Evaluation of the Work of Bronislaw Malinowski* (pp. 93-118). Routledge & Kegan Paul.
- Firth, J. R. C. (1968). A synopsis of linguistic theory. In F. Palmer (Ed.), *Selected Papers of J. R. Firth 1952-1959* (pp. 168-205). Longman.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. J. (1996). The search for units of meaning. *Textus*, 9(1), 75-106.
- 高歌, & 卫乃兴. (2019). 汉英翻译界面下的语义韵探究——来自《红楼梦》英译本的证据 [A semantic prosody study under the interface of Chinese-English translation: Evidence from the English versions of *Honglouloumeng*]. *解放军外国语学院学报* [Journal of PLA University of Foreign Languages], (1), 48-56.
- 高歌, & 卫乃兴. (2020). 英国语言学传统下的意义研究——从 Firth、Halliday 到 Sinclair [Meaning in the British linguistic tradition: From Firth, Halliday to Sinclair]. *中国外语* [Foreign Languages in China], (1), 25-34.
- 濮建忠. (2020). 扩展意义单位模型再解读 [Reinterpretation of the model of

- extended units of meaning]. 外语研究 [Foreign Languages Research], (2), 1-8+112.
- 王均松, & 田建国. (2016). 基于扩展意义单位模型的量词语义韵研究 [A corpus-based study on the semantic prosody of Chinese classifiers: Using the model of extended units of meaning]. 外语教学 [Foreign Language Education], (4), 39-43.
- 卫乃兴. (2002). 语义韵研究的一般方法 [Approaches to semantic prosody studies]. 外语教学与研究 [Foreign Language Teaching and Research], (4), 300-307.
- 卫乃兴. (2012). 共选理论与语料库驱动的短语单位研究 [The theory of co-selection and corpus-driven approach to phraseological units]. 解放军外国语学院学报 [Journal of PLA University of Foreign Languages], (1), 1-6+74+125.
- 周雅如, & 张奕. (2025). 中外学者语言学学术语篇自称语 we 的对比研究——基于扩展意义单位模型 [A contrastive study of self-mention “we” in linguistic scholarly discourse between Chinese and international scholars: Based on the extended unit of meaning model]. 唐山师范学院学报 [Journal of Tangshan Normal University], (2), 37-44.