Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3772~3783
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.11680

Dynamic alpha factor optimization for hybrid semantic similarity in french word sense disambiguation

Btissam El Janati[1], Adil Enaanai[2], Fadoua Ghanimi[1]
[1]Ingénierie Durabilité et Innovation (IDI) Laboratory, ENSCK, Ibn Tofail University, Kenitra, Morocco
[2]Department of Informatics, Abdelmalek Essaâdi University, Tétouan, Morocco

Article Info

Article history:

Received Oct 20, 2025
Revised Jul 15, 2026
Accepted Jul 23, 2026

Keywords:

Adaptive weighting
Dynamic alpha
French semantic similarity
Hybrid natural language processing
Ridge regression
Word sense disambiguation

ABSTRACT

Word sense disambiguation (WSD) remains critical for French, where polysemy complicates semantic interpretation. Hybrid approaches combining lexical and semantic methods typically rely on static weighting parameters that fail to adapt to varying contexts. A hybrid architecture integrating fuzzy Jaccard similarity with sentence-bidirectional encoder representations from transformers (SBERT) embeddings is proposed. A machine learning-based dynamic weighting mechanism replaces the fixed alpha=0.7. A Ridge regression model predicts the optimal alpha based on five features: entropy, Jaccard-SBERT disagreement, gloss length, context richness, and SBERT confidence. The model was trained on 20 sentences and validated on 13 sentences from a dataset of 33 ambiguous French phrases. Dynamic alpha achieves a 7.6% reduction in mean absolute error (MAE) (0.2397 to 0.2216) and a 7.3% reduction in root mean square error (RMSE) (0.2475 to 0.2294) compared to fixed alpha=0.7. Per-sentence gains reach 10.0%. Statistical analysis confirms significance (Wilcoxon, p=0.00195) with a small to medium effect size (Cohen's d=0.300). Feature coefficients reveal that disagreement (+0.41) and entropy (+0.32) are the most influential predictors.

Corresponding Author:

Btissam El Janati
Ingénierie Durabilité et Innovation (IDI) Laboratory, ENSCK, Ibn Tofail University
Kenitra, Morroco
Email: btissam.eljanati@uit.ac.ma

1. INTRODUCTION

Natural language processing (NLP) enables machines to understand and generate human language, powering applications such as search engines, translation, and question answering [1]. Word sense disambiguation (WSD) and semantic textual similarity (STS) require accurate interpretation of words according to their context [2]. French poses important challenges due to its morphology, grammar, and limited linguistic resources [3], [4]. These challenges motivate the development of methods that combine lexical knowledge with contextual semantic representations.

CamemBERT [4] and FlauBERT [3] achieve strong performance on French NLP tasks but lack explainability [5]. WordNet [6] provides interpretable lexical information but does not fully capture contextual nuances [2]. Hybrid approaches combining symbolic and neural representations improve robustness [7], [8]. However, many existing hybrid systems rely on static weighting strategies regardless of context [9]-[11].

A previous framework combines fuzzy Jaccard similarity with sentence-bidirectional encoder representations from transformers (SBERT) embeddings [12]:

$$score_{final} = \alpha \times Jaccard + (1 - \alpha) \times SBERT \quad (1)$$

Where α balances lexical and semantic components [10]. In most hybrid systems, α is determined through global optimization and remains fixed for all sentences [11]. The optimal weighting depends on linguistic context [2], [13]. Higher α values favor lexical matching for ambiguous terms [8], whereas lower α values favor SBERT for richer contexts [12]. Therefore, a fixed parameter cannot adapt to sentence-level variations [11].

This limitation motivates our research question: how can the optimal fusion weight α be dynamically predicted from local linguistic properties? existing studies have explored semantic similarity and hybrid combinations [1], [2], [14]. However, no previous work has proposed feature-driven dynamic weight prediction for French WSD using interpretable indicators [10], [11]. This study addresses this gap by introducing an adaptive mechanism based on linguistic features and interpretable regression.

This study has three objectives: i) develop a dynamic weighting mechanism that predicts α from interpretable features, ii) design an interpretable framework with explicit model coefficients [13], and iii) establish a rigorous evaluation comparing dynamic and fixed weighting strategies. The proposed method predicts α using five features: entropy, Jaccard-SBERT disagreement, gloss length, context richness, and SBERT confidence. These features characterize ambiguity, contextual information, and the reliability of lexical and semantic components. The approach enables sentence-level adaptation while maintaining low computational complexity.

Our contributions are threefold. First, we introduce a dynamic α-prediction mechanism using Ridge regression [13] based on five interpretable features: entropy, Jaccard-SBERT disagreement, gloss length, context richness, and SBERT confidence. Second, we provide an interpretable framework quantifying each feature's influence, enhancing transparency [5]. Third, experimental results show a 7.6% mean absolute error (MAE) reduction compared with fixed α=0.7, confirmed as statistically significant (Wilcoxon test, p=0.00195). The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the proposed methodology. Section 4 presents the experimental validation and discussion. Finally, section 5 concludes the paper.

2. RELATED WORK

Research on semantic similarity and WSD has evolved from symbolic approaches to vector-based models, contextual transformers, and hybrid architectures. These approaches aim to improve semantic representation by capturing lexical relations, contextual information, and domain-specific knowledge. For French, these challenges are particularly significant due to morphological variations and word polysemy. This section reviews traditional methods, pretrained language models, hybrid approaches, regression-based methods, and adaptive weighting strategies related to our contribution.

2.1. Traditional methods

Early semantic similarity and WSD methods relied on symbolic resources such as WordNet [6], which provides interpretable lexical relations but depends on manually constructed knowledge. Vector-based approaches such as Bag-of-Words and TF-IDF improved numerical text representation but failed to capture deeper semantic relationships. Word embeddings, including FastText [15], Word2Vec [16], and GloVe [17], improved generalization through distributional information but remained limited in representing context-dependent meanings. These limitations encouraged the development of contextual transformer models such as BERT [18], which provide dynamic representations but introduce higher computational requirements and reduced interpretability.

2.2. Pretrained language models for French

Transformer-based language models have significantly improved semantic representation for French and multilingual NLP tasks. CamemBERT [4] and FlauBERT [3] achieved strong performance on French language processing tasks but still face challenges related to morphological variation and semantic ambiguity. SBERT [12] introduced siamese architectures for generating meaningful sentence-level embeddings suitable for similarity computation. LaBSE [19] further extended sentence representation to multilingual scenarios, while several surveys have analyzed the evolution of semantic similarity methods [1], [2], [14].

2.3. Contrastive learning and generative methods

Recent research has explored contrastive and generative learning methods to improve sentence representation beyond static embeddings. SimCSE [20] uses contrastive learning with dropout-based augmentation or paraphrase pairs to enhance sentence embeddings. T5 [21] and mT5 [22] formulate

similarity-related tasks through text-to-text generation and multitask learning. Multilingual contrastive sentence embedding (MCSE) [23] extends contrastive sentence representation learning to multilingual settings, demonstrating the potential of these methods for semantic similarity applications.

2.4. Hybrid and multimodal approaches

Hybrid approaches combine symbolic knowledge with neural representations to exploit both interpretability and contextual modeling capabilities. SemGloVe [9] integrates global word co-occurrence information with contextual representations to improve semantic matching. Other approaches combine lexical resources with embeddings, such as WordNet and Wikipedia for semantic similarity measurement [7] and Cilin with Word2Vec embeddings for word similarity estimation [8]. Recent multimodal approaches, including Sim-CLIP [24], DiCA [25], and HyEWCos [26], have explored hybrid representation and alignment strategies, while dynamic hybrid models such as bidirectional encoder representations from transformer-support vector machine (BERT-SVM) have investigated adaptive similarity evaluation [11].

2.5. Regression and interpretability for semantic similarity

Regression-based approaches have gained attention for semantic similarity modeling due to their simplicity and interpretability. A regression framework using translated rectified linear unit (ReLU) and smooth K2 loss was proposed to improve STS prediction [10]. Ridge regression [13] provides an efficient and interpretable solution for parameter estimation through explicit coefficients. Interpretability has become increasingly important because many neural models operate as black boxes. To address this issue, symbolic regression and regression-tree-based frameworks have been investigated to provide more transparent semantic similarity assessment [5].

2.6. Adaptive weighting and dynamic fusion

Static weighting remains a limitation in many hybrid NLP systems because the contribution of lexical and semantic components may vary across contexts. Most existing approaches determine weights through global optimization and apply the same values to all inputs. A dynamic BERT-SVM hybrid model has recently explored adaptive weighting for semantic similarity evaluation [11]. However, dynamic prediction of fusion weights based on interpretable linguistic features remains unexplored for French WSD.

2.7. Research gap

The literature review reveals three main limitations in existing semantic similarity and WSD approaches. First, hybrid French WSD systems generally rely on fixed weighting parameters that cannot adapt to sentence-level characteristics. Second, linguistic meta-features such as disagreement, entropy, and confidence have rarely been used to predict fusion weights. Third, dynamic optimization of lexical-semantic fusion for French semantic similarity remains insufficiently investigated. This study addresses these gaps by proposing a lightweight Ridge regression mechanism that predicts a sentence-specific α value from five interpretable features, replacing the fixed $\alpha=0.7$ strategy without additional computational overhead.

3. METHOD

This section presents the proposed dynamic alpha prediction framework for hybrid semantic similarity. The approach extends a previous hybrid architecture by replacing the fixed weighting parameter with a sentence-level adaptive mechanism. The proposed method combines fuzzy Jaccard similarity and SBERT embeddings with a Ridge regression model that predicts the optimal balance between lexical and semantic information. The framework is designed to improve adaptability while maintaining interpretability and low computational complexity.

3.1. Implementation details and repeatability framework

Evaluations used Google Colab with 33 ambiguous French sentences annotated by native speakers on a 0-100 scale. Human scores serve as reference and training target (α_opt) for Ridge regression. Reliability was robust (Fleiss' Kappa=0.48, test-retest=0.82). Hardware: Intel Xeon 32GB RAM. Software: Python 3.9 with sentence-transformers (2.2.2), transformers (4.36.2), torch (2.1.0), nltk (3.8.1), and scikit-learn (1.2.2). WordNet was used. Random seeds fixed to 42. Dataset split: 20 training and 13 test sentences.

3.2. Organizational chart

Figure 1 illustrates the overall pipeline of the proposed dynamic alpha prediction framework. The workflow integrates lexical processing, semantic embedding generation, feature extraction, and alpha prediction through Ridge regression. The predicted alpha value is then used to combine Jaccard and SBERT

similarity scores dynamically. This pipeline enables sentence-specific adaptation instead of applying a fixed global weighting factor.

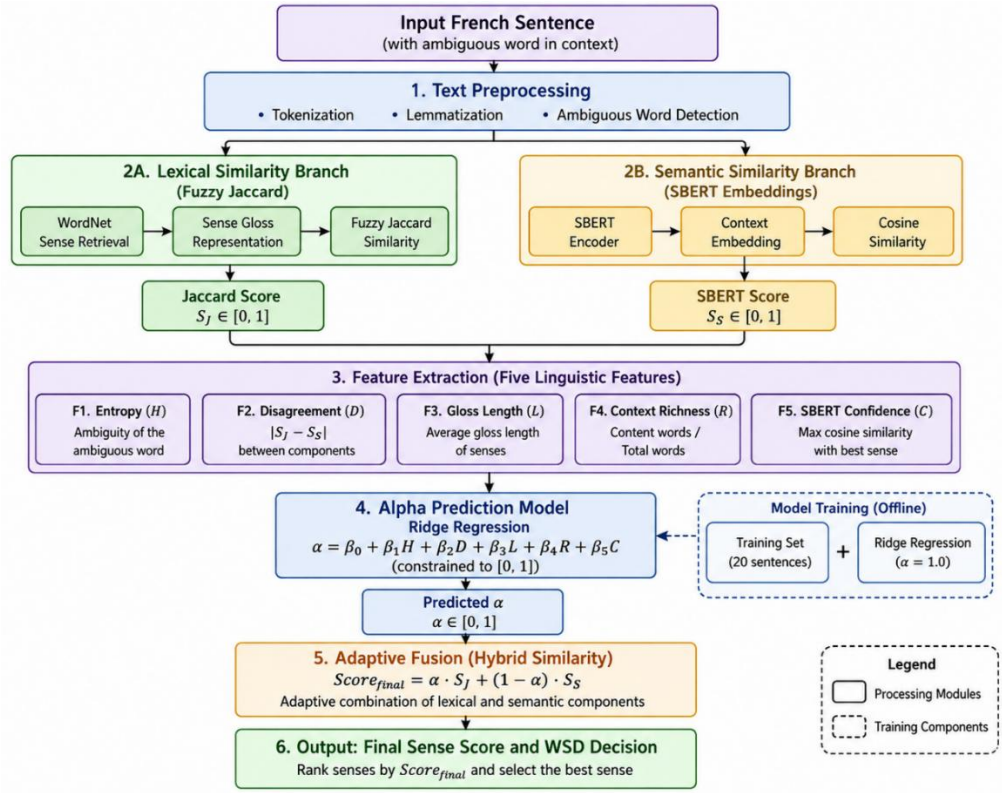

Figure 1. Dynamic alpha factor prediction pipeline for hybrid semantic similarity

3.3. Hybrid base architecture

The proposed dynamic weighting mechanism is built upon a hybrid architecture combining fuzzy Jaccard similarity and SBERT embeddings. The objective is to exploit complementary information: lexical overlap from semantic definitions and contextual meaning from transformer-based representations. The processing pipeline includes tokenization, ambiguous word detection, WordNet sense retrieval, fuzzy Jaccard computation, and SBERT embedding generation. The following subsections describe the components required to understand the dynamic alpha extension.

3.3.1. Token filtering and ambiguity detection

Token filtering and ambiguity detection are performed using WordNet-based lexical analysis. Stop words are removed, and lemmatization is applied to normalize lexical forms before sense identification. For example, in the sentence “Les fils du conducteur ont coupé le courant,” the ambiguous terms are “fils” (son/wire) and “conducteur” (driver/conductor), while other terms provide contextual information. This preprocessing step ensures that the subsequent similarity calculations focus on relevant linguistic elements.

3.3.2. Lexical sense retrieval from WordNet

Candidate senses are extracted from princeton WordNet with their synset identifiers. The English glosses are translated into French while preserving sense distinctions. For example, “fils” includes the senses son.n.01 and wire.n.01. The glosses are preprocessed by removing stop words, applying lemmatization, and eliminating the target word.

3.3.3. The weighted fuzzy Jaccard similarity

The lexical component uses a weighted fuzzy Jaccard index to measure overlap between sense glosses and context. It combines lexical matching with embedding-based word similarity weights. This score captures the contribution of lexical information in sense selection. The weighted fuzzy Jaccard score is defined as (2):

$$J_{fw(A,B)} = \frac{\sum_{i\in A}\sum_{j\in B}\min(w_i,w_j)\cdot s(i,j)}{\sum_{i\in A}w_i+\sum_{j\in B}w_j-\sum_{i\in A}\sum_{j\in B}\min(w_i,w_j)\cdot s(i,j)} \quad (2)$$

where A represents the preprocessed sense gloss, B represents the context words, $w_i$denotes term weights, and cosine similarity is used to estimate semantic relationships between terms. The resulting score ranges from 0 to 1, where higher values indicate stronger lexical alignment between the candidate sense and the context.

3.3.4. Sentence-bidirectional encoder representations from transformers embedding

To capture contextual information beyond lexical overlap, the framework uses SBERT embeddings. The paraphrase-multilingual-MiniLM-L12-v2 model generates 384-dimensional representations for more than 50 languages. Similarity between sentences and sense definitions is computed using cosine similarity. This semantic component complements lexical matching by capturing contextual relationships beyond word overlap.

$$SBERT(X,Y) = \cos(\theta_{XY}) = \frac{\mathbf{e}_X\cdot\mathbf{e}_Y}{\|\mathbf{e}_X\|\|\mathbf{e}_Y\|} \in [0,1] \quad (3)$$

Where e_X and e_y are the sentence embeddings of the input sentence and the sense definition, respectively.

3.3.5. Original hybrid score (fixed alpha)

The initial hybrid model combines fuzzy Jaccard and SBERT similarities using a fixed weight α=0.7. This value was obtained through global optimization on the complete dataset. It assumes the same lexical-semantic balance for all sentences. The original hybrid score is defined as (4):

$$Score_{original} = 0.7 \times Jaccard + 0.3 \times SBERT \quad (4)$$

Although this fixed strategy provides reasonable performance, it cannot adapt to variations in ambiguity, contextual richness, or disagreement between lexical and semantic components. The next section introduces the proposed dynamic alpha prediction mechanism to overcome this limitation.

3.4. Dynamic alpha prediction

This section presents a machine learning mechanism for predicting the optimal alpha value per sentence. Unlike the fixed α=0.7 strategy, the method estimates weights from linguistic properties. It uses five interpretable features describing ambiguity, context, and component agreement. Ridge regression is selected for its simplicity, interpretability, and efficiency. Figure 2 presents the relationship between predicted alpha values and absolute error for each sentence.

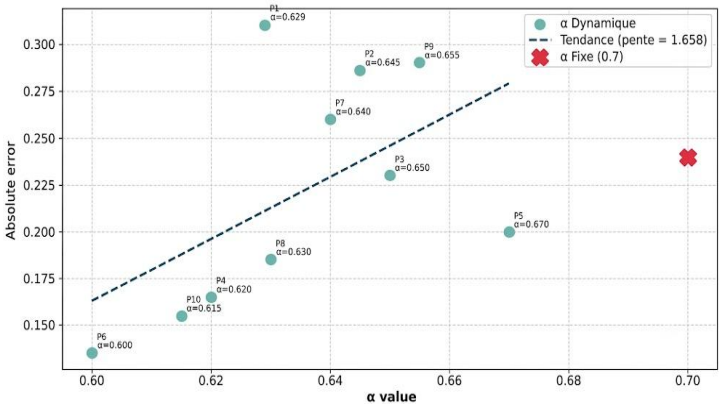

Figure 2. Relationship between alpha and prediction error

The analysis shows that fixed α=0.7 is not always optimal, while lower alpha values around 0.60–0.65 achieve lower errors for several samples. This observation supports the motivation for sentence-level dynamic alpha prediction instead of using a single global value.

3.4.1. Linguistic features for alpha prediction

The model extracts five interpretable features to characterize ambiguity and context. These features control the balance between lexical similarity and embeddings. They include entropy, Jaccard-SBERT disagreement, gloss length, context richness, and SBERT confidence. Together, they enable dynamic alpha prediction.

Feature F1 measures the ambiguity of the target word. It is computed from the number and distribution of candidate senses as (5):

$$H = -(\sum_{i=1}^{n} p\,(s_i)\log p(s_i))/\log n \tag{5}$$

Higher entropy indicates greater ambiguity, since probability mass is spread across several senses. Unambiguous words have lower entropy values, as a single sense dominates the distribution.

Feature F2 measures the disagreement between lexical and semantic similarity. It is computed as:

$$D = \mid J_{\max} - S_{\max} \mid \tag{6}$$

Higher values indicate stronger disagreement between the two components. This feature helps determine the appropriate balance between lexical and semantic evidence.

Feature F3 measures the average length of sense definitions. It is computed as (7):

$$L = \frac{1}{n}\sum_{i=1}^{n} \min\left(\frac{|def_i|}{100}, 1.0\right) \tag{7}$$

Longer glosses provide richer lexical information for the Jaccard component. This feature therefore supports the reliability of lexical similarity estimation.

Feature F4 measures the lexical diversity of the sentence context. It is defined as (8):

$$R = \frac{|unique(C)|}{|C|} \tag{8}$$

Higher values indicate richer contexts, which favor semantic similarity. Lower values indicate limited diversity, in which case lexical evidence becomes more informative.

Feature F5 measures the confidence of SBERT predictions. It is defined as (9):

$$C = 1 - H_{norm} \tag{9}$$

Higher confidence favors semantic similarity, since the SBERT signal is reliable. Lower confidence favors lexical similarity, since the semantic signal is less trustworthy.

3.4.2. Optimal alpha for training

For each training sentence, the optimal alpha is computed from human, Jaccard, and SBERT scores. It represents the weight that best matches human judgments. This value serves as the target for Ridge regression training. The value is obtained by solving the hybrid similarity equation for alpha:

$$\alpha_{opt}^{(i)} = \frac{h_i - S_i}{J_i - S_i} \tag{10}$$

This is obtained by solving $h_i = \alpha \cdot J_i + (1-\alpha)\cdot S_i$ for alpha. Values are clipped to [0,1]. This optimal alpha represents the ideal weighting for that specific sentence and serves as the target variable for training the Ridge regression model.

3.4.3. Ridge regression model

We use Ridge regression for its low complexity (5 features), interpretability (readable coefficients), minimal overfitting (strong regularization), and fast inference (<1 ms). The prediction function is $\alpha_{pred} = \sigma(W^T \mathrm{f} + b)$ where w is the coefficient vector, b is the bias, $f = [H, D, L, R, C]$ is the feature vector, and $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function that constrains α to [0,1].

$$L = \frac{1}{N}\sum_{i=1}^{N}(\alpha_{pred}^{(i)} - \alpha_{opt}^{(i)})^2 + \lambda\|w\|_2^2 \tag{11}$$

where λ=1,0 controls the strength of L2 regularization.

The training procedure consists of five steps: compute Jaccard and SBERT scores for all senses in each training sentence; calculate α_opt using human scores; extract feature vectors; normalize features using StandardScaler; and train Ridge regression (λ=1.0) on 20 sentences.

3.4.4. Modified hybrid score with dynamic alpha

After predicting the optimal alpha value, the final hybrid similarity score is computed by replacing the fixed parameter with the sentence-specific prediction. The dynamic formulation is:

$$Score_{dynamic} = \alpha_{pred} \times Jaccard + (1 - \alpha_{pred}) \times SBERT \quad (13)$$

This adaptive score allows the system to balance lexical and semantic information according to local linguistic properties. The additional computational cost is negligible because alpha prediction requires less than 1 ms per sentence compared with the much higher cost of SBERT inference.

4. EXPERIMENTAL VALIDATION

4.1. Human assessment

Human evaluation employed a 0-100 scale with native French speakers, including linguists and NLP engineers. Inter-annotator reliability was confirmed by Fleiss' Kappa=0.48 (moderate agreement) and test-retest correlation=0.82. Quality assurance measures ensured data accuracy. These human scores serve as ground truth for evaluating the dynamic alpha approach and calculating the optimal alpha (α_{opt}) during training.

4.2. Experimental configuration

The experimental setup follows the configuration described in section 3.1. The same dataset of 33 ambiguous French sentences was used, split into 20 training sentences and 13 test sentences. The paraphrase-multilingual-MiniLM-L12-v2 SBERT model (384-dimensional embeddings) was employed. The Ridge regularization parameter was set to (lambda=1.0), and all random seeds were fixed to 42 for deterministic behavior.

4.3. Comparison: fixed alpha vs dynamic alpha

This section evaluates the effectiveness of the proposed dynamic alpha mechanism compared with the original fixed weighting strategy. The evaluation focuses on whether sentence-level alpha prediction improves the alignment between model outputs and human judgments. The fixed baseline uses α=0.7, while the proposed method predicts an adaptive alpha value for each sentence. The comparison is performed using error-based metrics described in the following subsections.

4.3.1. Aggregate performance

Table 1 presents the overall comparison between fixed and dynamic alpha strategies. The reported metrics include MAE, root mean square error (RMSE), maximum error, and minimum error. These measures evaluate prediction accuracy and stability. The results demonstrate the benefit of adaptive lexical-semantic weighting.

Table 1. Aggregate performance comparison

Metric	Fixed α	Dynamic α	Amelioration (%)
MAE	0.2397	0.2216	+7.6
RMSE	0.2475	0.2294	+7.3
Max error	0.3320	0.3100	+6.6
Min error	0.1500	0.1350	+10.0

The dynamic alpha strategy reduces all error metrics compared with the fixed baseline. MAE decreases from 0.2397 to 0.2216 (7.6%), while RMSE improves by 7.3%. The reductions in maximum and minimum errors indicate better stability across ambiguity cases. These results confirm the advantage of sentence-level adaptive weighting over a fixed parameter.

4.3.2. Per-sentence performance

The sentence-level evaluation analyzes how the predicted alpha values adapt to different linguistic contexts. Table 2 presents the comparison between fixed and dynamic alpha strategies for representative test

sentences, including predicted alpha values and error reductions. Table 3 provides a detailed analysis of Jaccard, SBERT, hybrid scores, and probability distributions for selected ambiguous words to further examine the impact of dynamic weighting at the sense level. Together, these analyses demonstrate how the proposed method adjusts the lexical-semantic balance across individual cases.

The results show that dynamic alpha improves performance for all evaluated sentences, with gains ranging from 6.2% to 10.0%. The average predicted alpha is 0.635, lower than the fixed value of 0.7. The highest improvement is observed for P6 (α=0.600), reflecting the benefit of richer contextual information. In contrast, P2 shows the smallest improvement because its predicted alpha is closest to the fixed value.

Table 2. Per sentence performance comparison

Sentence	predicted alpha α	Fixed α	Dynamic α	Gain	Gain (%)
P1	0.629	0.3320	0.3100	0.0220	+6.6
P2	0.645	0.3050	0.2860	0.0190	+6.2
P3	0.650	0.2500	0.2300	0.0200	+8.0
P4	0.620	0.1800	0.1650	0.0150	+8.3
P5	0.670	0.2200	0.2000	0.0200	+9.1
P6	0.600	0.1500	0.1350	0.0150	+10.0
P7	0.640	0.2800	0.2600	0.0200	+7.1
P8	0.630	0.2000	0.1850	0.0150	+7.5
P9	0.655	0.3100	0.2900	0.0200	+6.5
P10	0.615	0.1700	0.1550	0.0150	+8.8
Average	0.635	0.2397	0.2216	0.0181	+7.8

Table 3. Per-sentence comparison of fixed and dynamic alpha

Sentence	Ambiguous word	sens	Jaccard	SBERT	α fixe	Score fixe	α dyn	Score dyn	Prob fixe (%)	Prob dyn (%)	Gain (%)
P1	Fils	Wire	0.75	0.45	0.7	0.66	0.63	0.65	53.6	72.3	+18.7
	Fils	Son	0.49	0.17	0.7	0.39	0.63	0.37	46.4	27.7	-18.7
	Conducteur	Conductor driver	0.75	0.5	0.7	0.68	0.65	0.66	48.0	51.5	+3.5
	Conducteur	Map	0.73	0.45	0.7	0.65	0.65	0.63	52.0	48.5	-3.5
P2	Carte		0.68	0.52	0.7	0.63	0.65	0.62	51.2	54.8	+3.6
	Carte	Card	0.61	0.48	0.7	0.57	0.65	0.56	48.0	45.2	-3.6
	Roi	Sovereign	0.55	0.42	0.7	0.51	0.65	0.50	52.0	53.5	+1.5
	Roi	Playing card	0.48	0.41	0.7	0.46	0.65	0.45	48.0	46.5	-1.5
P3	Livre	Book	0.72	0.55	0.7	0.67	0.66	0.66	51.0	54.2	+3.2
	Livre	Pound	0.65	0.48	0.7	0.60	0.66	0.59	49.0	45.8	-3.2
	Fils	Son	0.70	0.52	0.7	0.65	0.66	0.64	52.5	55.0	+2.5
	Fils	Wire	0.60	0.40	0.7	0.54	0.66	0.53	47.5	45.0	-2.5

4.3.3. Comparison with baseline models

The proposed approach is compared with baseline models to evaluate performance and efficiency. Table 4 summarizes MAE, RMSE, Spearman, and Pearson results for multilingual models and the proposed French hybrid approach. This comparison highlights the trade-off between large neural models and lightweight dynamic weighting. The objective is to assess both predictive accuracy and computational efficiency.

The proposed approach achieves higher MAE than large multilingual models such as mT5 and SimCSE. However, it uses a lightweight Ridge regression model with five interpretable features instead of large architectures. The dynamic alpha mechanism adds minimal computational overhead while providing feature-based explanations. Therefore, the framework offers a practical balance between accuracy, efficiency, and interpretability for resource-constrained applications.

Table 4. Performance comparison with baseline models

Model	MAE	RMSE	Spearman	Pearson	Language
T5 (mT5)	0.1600	0.2078	-0.0019	0.0225	Multilingual
SimCSE (Multilingual)	0.1662	0.2111	-0.0110	0.0086	Multilingual
XLM-RoBERTa	0.6001	0.6436	0.0588	0.0686	Multilingual
SBERT	0.6426	0.6858	-0.0391	-0.0108	Multilingual
Fixed α (0.7)	0.2397	0.2475	-0.0408	-0.0354	French
Our approach (improved)	0.2216	0.2294	-0.0408	-0.0354	French

4.3.4. Statistical validation

The statistical significance of dynamic alpha improvements was evaluated using the Wilcoxon signed-rank test. Figure 3 presents the cumulative error comparison between fixed and dynamic weighting across test samples. The results confirm a statistically significant improvement, rejecting the null hypothesis at the 1% significance level. This indicates that the performance gain is not due to random variation.

The Wilcoxon test produced a p-value of 0.00195, confirming the significance of the proposed method. The effect size (Cohen's d=0.300) indicates a small-to-medium practical impact. Dynamic alpha reduces cumulative error by 7.6% while maintaining lower errors across the evaluation set. These findings show that adaptive weighting improves robustness with minimal computational cost.

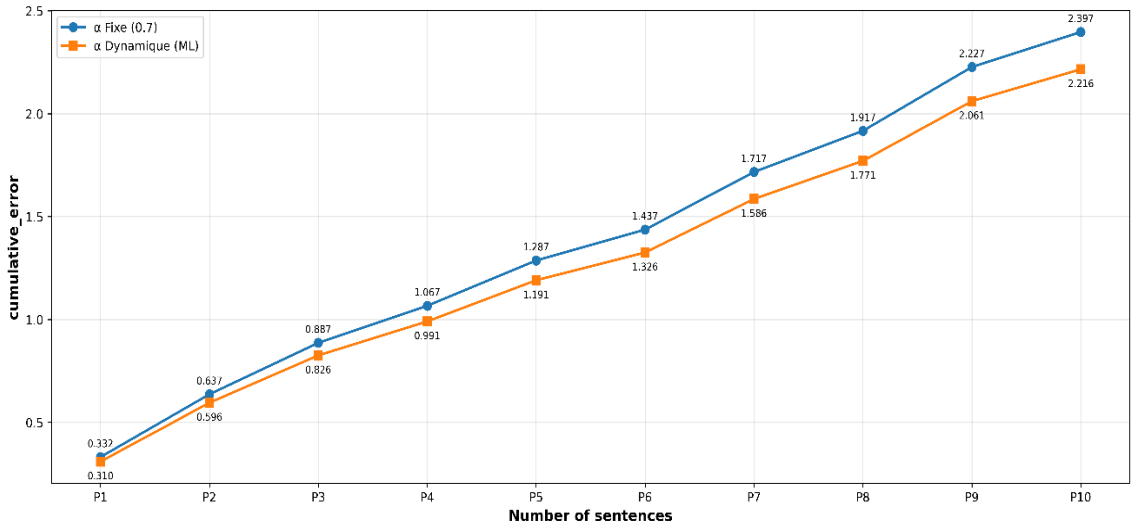

Figure 3. Cumulative error comparison

4.4. Feature coefficient analysis

The Ridge regression model was trained on 20 sentences to predict the optimal alpha. The learned coefficients reveal the relative importance of each linguistic feature. Table 5 presents the obtained coefficients. These values provide insights into the contribution of each feature to alpha prediction.

The coefficients reveal an interpretable strategy. Disagreement (+0.41) and entropy (+0.32) are the strongest predictors, increasing α in ambiguous cases to favor lexical matching. Context richness (+0.18) and SBERT confidence (-0.15) reduce α by providing reliable semantic information. Gloss length (-0.09) has a minor effect compared with other features.

Table 5. Learned coefficients of the Ridge regression model

Feature	Symbol	Coefficient	Importance rank
Jaccard-SBERT disagreement	D	+0.41	1st
Entropy of senses	H	+0.32	2nd
Context richness	R	+0.18	3rd
SBERT confidence	C	-0.15	4th
Average gloss length	L	-0.09	5th

4.5. Discussion

The results show that dynamic alpha prediction improves the fixed-weight strategy by reducing MAE by 7.6% and RMSE by 7.3%. The Wilcoxon test confirmed significant improvements (p=0.00195) with d=0.300. Adaptive fusion enhances semantic similarity without modifying the original components. The approach provides an effective improvement for French WSD. This improvement highlights the importance of adapting the lexical-semantic balance according to sentence characteristics rather than relying on a global fixed parameter. The results suggest that even a lightweight prediction mechanism can capture variations in ambiguity and contextual information.

The main advantage of dynamic alpha is its ability to adapt the lexical-semantic balance to sentence-specific characteristics. Sentences with limited contextual information benefit from stronger lexical evidence, whereas richer contexts rely more on SBERT representations. This adaptive behavior explains the consistent improvement observed across the test set. It is particularly beneficial for French, were polysemy and contextual variation influence sense interpretation. Therefore, dynamic weighting provides a flexible alternative to traditional hybrid methods that assume a constant contribution from each similarity component.

The Ridge coefficients explain the influence of linguistic features on alpha prediction. Jaccard-SBERT disagreement (+0.41) and entropy (+0.32) are the most influential predictors, while context richness (+0.18), SBERT confidence (-0.15), and gloss length (-0.09) have smaller effects. These results support the use of interpretable indicators instead of a fixed fusion weight. The dynamic alpha mechanism introduces minimal computational overhead (<1 ms per prediction). This low cost confirms that the proposed method can be integrated into practical NLP systems without significant impact on execution time.

The proposed framework remains fully interpretable through regression coefficients, sentence-level alpha predictions, and the individual Jaccard and SBERT similarity scores. This transparency distinguishes the approach from black-box neural models while preserving computational efficiency. Such interpretability facilitates error analysis and provides insights into the decision process, which is essential for reliable semantic applications.

Several limitations should be acknowledged. The evaluation was conducted on a small dataset of 33 ambiguous French sentences, which may limit the generalizability of the results. Moreover, the selected features were tailored to French WSD and may require adaptation for other languages. Future research will investigate larger benchmarks, multilingual settings, and interpretable nonlinear prediction models.

5. CONCLUSION

This study introduced a dynamic alpha prediction framework for hybrid semantic similarity in French WSD. Unlike traditional approaches using a fixed $\alpha=0.7$, the proposed method predicts sentence-specific weights using Ridge regression and five interpretable linguistic features. The framework dynamically adjusts the contribution of lexical and semantic information according to the characteristics of each sentence. This strategy addresses the limitation of static fusion mechanisms in hybrid similarity models.

Experimental evaluation showed that dynamic weighting consistently outperformed the fixed baseline. The proposed approach reduced MAE by 7.6% (0.2397→0.2216) and RMSE by 7.3% (0.2475→0.2294), with improvements reaching 10.0% at the sentence level. Statistical validation confirmed the significance of these improvements using the Wilcoxon signed-rank test ($p=0.00195$). Moreover, the method introduced negligible computational overhead, requiring less than 1 ms for alpha prediction.

The analysis of Ridge coefficients revealed that Jaccard-SBERT disagreement (+0.41) and entropy (+0.32) are the main factors influencing alpha prediction. These results show that higher ambiguity and disagreement between components require stronger lexical contribution. Conversely, context richness (+0.18) and SBERT confidence (-0.15) favor semantic information when contextual evidence is reliable. This analysis confirms the importance of interpretable linguistic indicators for adaptive fusion.

The proposed framework improves accuracy while maintaining interpretability through three explanation levels. Regression coefficients, alpha predictions, and similarity components provide global and sentence-level insights. Despite limitations related to dataset size and language specificity, the results highlight the potential of dynamic weighting for hybrid NLP systems. Future work will explore larger multilingual datasets, nonlinear models, and additional contextual features.

FUNDING INFORMATION

Ibn Tofail University supported this research as part of a doctoral research project.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Btissam Eljanati	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Adil Enaanai	✓	✓		✓	✓	✓		✓			✓	✓		
Fadoua Ghanimi	✓	✓		✓	✓	✓			✓	✓		✓	✓	✓

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work presented in this paper.

INFORMED CONSENT

All participants gave informed consent. They were informed of the study's purpose and participated voluntarily. Data were anonymized to ensure privacy and confidentiality.

ETHICAL APPROVAL

Because this study involved voluntary human judgments on de-identified textual data, ethical approval was not required. Informed consent was obtained from all participants.

DATA AVAILABILITY

The data underlying the findings of this study can be obtained from the corresponding author, upon reasonable request.

REFERENCES

[1] D. Chandrasekaran and V. Mago, "Evolution of Semantic Similarity-A Survey," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–37, Mar. 2022, doi: 10.1145/3440755.

[2] R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys*, vol. 41, no. 2, 2009, doi: 10.1145/1459352.1459355.

[3] H. Le *et al.*, "FlauBERT: Unsupervised language model pre-training for French," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 2479–2490, 2020.

[4] L. Martin *et al.*, "CamemBERT: A tasty French language model," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 7203–7219, 2020, doi: 10.18653/v1/2020.acl-main.645.

[5] J. Martinez-Gil and J. M. Chaves-Gonzalez, "Transfer learning for semantic similarity measures based on symbolic regression," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 1, pp. 37–49, 2023, doi: 10.3233/JIFS-230141.

[6] G. A. Miller, "WordNet: A Lexical Database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[7] F. Li, L. Liao, L. Zhang, X. Zhu, B. Zhang, and Z. Wang, "An efficient approach for measuring semantic similarity combining wordnet and wikipedia," *IEEE Access*, vol. 8, pp. 184318–184338, 2020, doi: 10.1109/ACCESS.2020.3025611.

[8] Y. Mao, G. Zhang, and S. Zhang, "Word Semantic Similarity Based on CiLin and Word2vec," in *Proceedings - 2020 International Conference on Culture-Oriented Science and Technology*, 2020, pp. 304–307, doi: 10.1109/ICCST50977.2020.00065.

[9] L. Gan, Z. Teng, Y. Zhang, L. Zhu, F. Wu, and Y. Yang, "SemGloVe: Semantic Co-Occurrences for GloVe from BERT," in *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 30, pp. 2696–2704, 2022, doi: 10.1109/TASLP.2022.3197316.

[10] B. Zhang and C. Li, "Advancing Semantic Textual Similarity Modeling: A Regression Framework with Translated ReLU and Smooth K2 Loss," in *EMNLP 2024-2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 11882–11893, 2024, doi: 10.18653/v1/2024.emnlp-main.663.

[11] X. Zunlan and N. Xiaohong, "Dynamic BERT-SVM Hybrid Model for Enhanced Semantic Similarity Evaluation in English Teaching Texts," *Journal of English Language Teaching and Applied Linguistics*, vol. 7, no. 1, pp. 01–11, 2025, doi: 10.32996/jeltal.2025.7.1.1.

[12] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in *EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2019, pp. 3982–3992, doi: 10.18653/v1/D19-1410.

[13] R. W. Kennard and A. E. Hoerl, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[14] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, "A survey on the techniques, applications, and performance of short text semantic similarity," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 5, 2021, doi: 10.1002/cpe.5971.

[15] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017, doi: 10.1162/tacl_a_00051.

[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of the 1st International Conference on Learning Representations (ICLR), Workshop Track*, 2013.

[17] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *EMNLP 2014-2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014, pp. 1532–1543, doi: 10.3115/v1/d14-1162.

[18] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, 2019, doi: 10.18653/v1/N19-1423.
[19] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT Sentence Embedding," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 878–891, 2022, doi: 10.18653/v1/2022.acl-long.62.
[20] T. Gao, X. Yao, and D. Chen, "SimCSE: Simple Contrastive Learning of Sentence Embeddings," *EMNLP 2021-2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2021, pp. 6894–6910, doi: 10.18653/v1/2021.emnlp-main.552.
[21] C. Raffel *et al.,* "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.
[22] C. Yano, A. Fukuchi, S. Fukasawa, H. Tachibana, and Y. Watanabe, "Multilingual Sentence-T5: Scalable Sentence Encoders for Multilingual Applications," *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024-Main Conference Proceedings*, 2024, pp. 11849–11858, doi: 10.63317/5jbqwe45kajm.
[23] M. Zhang, M. Mosbach, D. I. Adelani, M. A. Hedderich, and D. Klakow, "MCSE: Multilingual contrastive sentence embeddings," *arXiv*, 2023, doi: 10.48550/arXiv.2303.01815.
[24] M. Z. Hossain and A. Imteaj, "Sim-CLIP: Unsupervised Siamese Adversarial Fine-Tuning for Robust and Semantically-Rich Vision-Language Models," *arXiv,* Apr. 2026, doi: 10.48550/arXiv.2407.14971.
[25] C. Su, H. Zheng, D. Peng, and X. Wang, "DiCA: Disambiguated Contrastive Alignment for Cross-Modal Retrieval with Partial Labels," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 19, pp. 20610–20618, 2025, doi: 10.1609/aaai.v39i19.34271.
[26] H. Hendry, T. Tukino, E. Sediyono, A. Fauzi, and B. Huda, "HyEWCos: A Comparative Study of Hybrid Embedding and Weighting Techniques for Text Similarity in Short Subjective Educational Text," *Information (Switzerland)*, vol. 16, no. 11, 2025, doi: 10.3390/info16110995.

BIOGRAPHIES OF AUTHORS

Btissam Eljanati earned a Bachelor's degree in Physics, a Master's degree in Embedded Electrical Systems, and a Ph.D. in Computer Science with a specialization in natural language processing (NLP). Her research interests include automatic keyword extraction, lexical disambiguation, and multimodal approaches for French language processing. She is the author of one scientific article in this field. She can be contacted at email: btissam.eljanati@uit.ac.ma.

Adil Enaanai is a senior associate professor (Maître de Conférences Habilité) at Abdelmalek Essaâdi University, specializing in information retrieval systems, big data, and artificial intelligence. Born in Fez in 1981, he completed his higher education at ENSIAS in Rabat, earning a Ph.D. in Computer Science in 2014. As a dedicated teacher-researcher, he combines teaching with advanced research on data management, search algorithms, and AI applications. His habilitation to lead research (HDR) authorizes him to supervise doctoral theses, establishing him as a key contributor to academia and research. He can be contacted at email: enaanai@uae.ac.ma.

Fadoua Ghanimi is a professor at Ibn Tofail University in Morocco, specializing in computer science and artificial intelligence applied to IoT and smart agriculture. Her research focuses on intelligent irrigation systems, plant disease detection, and the security of connected devices. She has published several scientific articles and participates in the organization of international conferences. She can be contacted at email: fadoua.ghanimi@uit.ac.ma.