A hybrid approach for measuring semantic similarity in lexically identical but ambiguous sentences

Btissam El Janati[1], Adil Enaanai[2], Fadoua Ghanimi[1]
[1]Department of Physics Engineering, Faculty of Science, Ibn Tofail University, Kenitra, Morocco
[2]Department of informatic, Abdelmalek Saadi University, Tétouan, Morocco

Article Info

ABSTRACT

This study addresses the critical challenge of semantic similarity and lexical disambiguation in natural language processing, focusing on sentences with structural and lexical ambiguities. We introduce an innovative hybrid approach that synergistically combines symbolic and neural methods to better align with human judgment. Our methodology dynamically integrates fuzzy Jaccard's lexical precision with SBERT embeddings' contextual sensitivity, enabling adaptive semantic ambiguity resolution. Experimental evaluation on 33 ambiguous sentences demonstrates that our approach significantly outperforms conventional artificial intelligence (AI) systems, achieving an 11.7% reduction in mean absolute error compared to reference models, with statistical analysis confirming robust results (d = -0.80, p < 0.001). This represents a 65% improvement in human evaluation alignment over existing methods. Our research contributes to advancing the field by showing that architectural intelligence can surpass mere parameter scaling, offering an effective solution for applications requiring both precision and interpretability, with promising directions for multilingual extension and explainable AI integration.

Corresponding Author:

Btissam El Janati
Department of Physics, Faculty of Science, Ibn Tofail University
Kenitra, Morroco
Email: Btissam.eljanati@uit.ac.ma

1. INTRODUCTION

Natural language processing (NLP) enables machines to interpret and generate human language, forming the foundation of applications such as search engines, machine translation, and dialogue systems. Among its core challenges, semantic textual similarity (STS) aims to quantify how closely two sentences convey the same meaning. While significant progress has been made in English NLP, French language processing faces unique challenges due to its rich morphology, grammatical complexity, and limited annotated resources.

Existing approaches for French semantic similarity reveal critical limitations. Neural models like CamemBERT and FlauBERT, while effective for general tasks, often struggle with fine-grained semantic distinctions and lack interpretability. Symbolic methods offer transparency but fail to capture contextual nuances. Most notably, current hybrid approaches for French either rely on static combinations of components or fail to provide adaptive disambiguation mechanisms, particularly for lexically identical but semantically divergent sentences. This research gap is especially pronounced in French, where existing methods cannot adequately handle the language's complex agreement rules and contextual dependencies.

This study addresses these limitations through three primary objectives:
a. Develop a dynamic hybrid framework that adaptively integrates symbolic and neural methods for French semantic similarity.
b. Create an interpretable disambiguation system that maintains contextual sensitivity while providing transparent decision-making.
c. Establish a robust evaluation methodology that accurately measures alignment with human semantic judgments.

This work makes three significant contributions to French NLP:
− An adaptive hybrid architecture that surpasses the limitations of static combination approaches.
− Demonstrated effectiveness on complex French ambiguities where current methods fail.
− A validation framework showing 11.7% mean absolute error (MAE) reduction and 65% better human alignment compared to state-of-the-art systems.

The remainder of this paper is organized as follows: section 2 reviews related work, section 3 details our methodology, section 4 presents experimental validation, section 5 discusses findings, and section 6 concludes with future directions.

2.   RELATED WORK

Research on semantic similarity and word sense disambiguation (WSD) has evolved from symbolic and rule-based systems to vector-based models and finally to deep neural and multimodal architectures. Each paradigm addresses semantic representation from different theoretical and computational perspectives, particularly for French, where morphological richness and polysemy make disambiguation especially challenging.

2.1. Classical approaches

Early methods for semantic similarity relied on symbolic and rule-based frameworks such as WordNet, which modeled syntactic and semantic relations through logical inference. These systems were highly interpretable but faced major challenges: manual rule dependence, poor scalability, and limited linguistic adaptability [1].

In Arabic NLP, for example, Enaanai *et al.* [2] proposed a morphosemantic filtering method addressing ambiguity caused by agglutination and lack of vocalization. Their system expanded words into all possible vocalized forms and disambiguated meanings using contextual rules, a user profile, and the semantic lexicon ARRAMOUZ ALWASEET, with a Jaccard-based similarity function to enhance alignment. Though effective, this approach remained constrained by rule complexity and limited portability to other languages.

Subsequent vector-based models, including bag-of-words (BoW) and TF-IDF [1], provided quantitative text representations but ignored semantic relationships. Word embedding models improved lexical generalization FastText [3], [4] introduced subword information, while Word2Vec [5] and GloVe [6] captured co-occurrence patterns. However, these models were context-insensitive and struggled with polysemy. Later probabilistic embeddings, such as Arora *et al.* [7], enhanced global structure but still relied on simplifying Gaussian assumptions.

These limitations motivated the emergence of contextualized transformer-based models such as BERT [8], which introduced dynamic embeddings sensitive to surrounding words, improving semantic coherence but raising issues of computational cost and interpretability.

2.2. Pretrained language models

Pretrained language models (PLMs) based on the transformer architecture have transformed NLP by enabling context-aware and multilingual representations. For French, notable models include CamemBERT [6], a RoBERTa-based model trained on OSCAR and French Wikipedia; FlauBERT [1], a large 12–24-layer model with rich lexical coverage; BERTweetFR [9]-[12], specialized for Twitter; and mBERT [8] a multilingual model less optimized for French syntax and morphology. Semantic similarity is usually computed from the [CLS] token or mean-pooled embeddings using cosine similarity. To better capture sentence-level meaning, Sentence-BERT (SBERT) introduced a Siamese architecture, later adapted for French in CamemBERT-Sentence and LaBSE [13]. Benchmarks such as STS-FR provide evaluation datasets, though they are limited in domain diversity. Despite their success, PLMs still face challenges with morphological variation, register diversity, and biases from limited French corpora, which can lead to semantic drift in polysemous contexts, highlighting the need for more robust and context-sensitive approaches.

2.3. Contrastive learning and supervised methods

Recent approaches have advanced beyond static embeddings by incorporating contrastive and generative learning frameworks. SimCSE [14] refines sentence representations through contrastive learning, using dropout-induced positives in unsupervised mode and paraphrase pairs with negatives in supervised mode. French adaptations fine-tuned on FrSemLex improved STS-FR performance and semantic precision in sentence alignment.

T5 [15] and its French variant ST5 [16], [17] reconceptualize similarity as a text generation task, producing reformulations or similarity scores through multitask learning. Empirical results show SimCSE [2] -RoBERTa-large (2021) reaching 86.7 on STS-B, while ST5 and mST5 achieve up to $\rho = 83.2$ on XSTS, confirming the complementarity between contrastive and generative paradigms for semantic similarity tasks.

2.4. Hybrid and multimodal approaches for advanced NLP

Hybrid and multimodal architectures combine symbolic interpretability with deep contextual understanding. Systems such as SemGloVe [18], [19] integrate GloVe's global co-occurrence statistics with BERT's contextual embeddings, achieving 12% gains in lexical similarity and improved WSD.

Cross-modal models like Sim-CLIP [20], DiCA [21], and MCSE [22] extend these principles to vision–language and audio–text domains, using contrastive alignment for partially labeled or unannotated datasets. Recent advances in multilingual multimodal modeling, such as multilingual Sentence-T5 [23], ensure cross-lingual semantic consistency and enable cross-modal disambiguation, benefiting multilingual translation and multimodal retrieval. Despite these achievements, most studies focus on English, with limited exploration in French especially for lexically identical but semantically divergent sentences, which remain a critical challenge for context-sensitive understanding.

2.5. Comparative summary of major model families

Table 1 summarizes major NLP model families, highlighting representative models, key advancements, limitations, and reported performance to provide a concise comparative overview of current approaches.

Table 1. NLP models families: features, limitations, and performances

Model family	Representative models	Key features/progress	Limitations	Performance/Notes	Model family
Vector-based	LSA, Word2Vec, GloVe, Arora	Efficient embeddings, latent/global semantics	No context, ignores word order, memory-heavy	–	Vector-based
Transformer PLMs	BERT, FlauBERT, SBERT, CamemBERT, BERTweetFR	Contextual embeddings, sentence-level pooling	High cost, low explainability	MNLI:84.6, STS-B:84.9, TextClf:95%, F1:71.27	Transformer PLMs
Contrastive / Generative	SimCSE, T5, m-ST5	Contrastive learning, generative modeling	Needs fine-tuning, data-dependent	STS-B:76.85–86.70, GLUE:90.3, XSTS ρ:83.2	Contrastive / Generative
Hybrid / Multimodal	SemGloVe, MCSE, DiCA, Sim-CLIP4	Embedding fusion, multimodal learning	Computationally heavy	Accuracy:95.68%, mAP:0.697, CIDEr:81.6	Hybrid / Multimodal

2.6. Discussion and identified research gap

Although French PLMs such as CamemBERT and FlauBERT have achieved notable progress, they still struggle with polysemy and semantic ambiguity at the sentence level. Existing benchmarks, such as STS-FR and FrSemLex, remain limited and lack examples where lexical identity hides divergent meanings (e.g., fils, courant, branche).

Moreover, hybrid frameworks combining symbolic interpretability and neural reasoning remain underexplored for French. Recent multilingual models, such as mDeBERTa [9], LLaMA [18], and XSIM23 [1], show promising potential for modeling cross-lingual semantic ambiguity, yet their effectiveness for contextual disambiguation in French remains to be investigated.

Our proposed hybrid method directly addresses lexically identical yet semantically ambiguous French sentences, bridging the gap between human-like interpretability and the contextual accuracy of deep learning representations. The novelty of our approach lies in its context-aware fusion mechanism, which dynamically adjusts the interplay between symbolic and neural elements according to linguistic complexity. In contrast to traditional hybrid models relying on fixed weightings, our framework continuously assesses ambiguity cues to fine-tune the trade-off between precision and transparency. This signifies a transformative shift from rigid architectures to an adaptive system tailored to the subtleties of French, effectively tackling persistent issues like homonymy and structural ambiguity that challenge current PLMs and hybrid solutions.

3. METHOD

3.1. Instruments

Our hybrid approach uses the sentence-transformers/all-mpnet-base-v2 version of SBERT, generating 768-dimensional embeddings. Experiments were conducted on Google Colab in a CPU environment with the following parameters: batch size of 32, temperature of 0.05 for the SoftMax function, and similarity threshold of 0.6. The implementation integrates WordNet for synset extraction, NLTK and SpaCy for linguistic preprocessing, and custom routines for calculating weighted fuzzy Jaccard. The fusion mechanism dynamically combines cosine similarities (with L2 normalization) and symbolic similarities through an adaptive attention function based on detected linguistic complexity.

3.2. Organigramme

The hybrid method combines linguistic disambiguation and semantic vector modeling through eight sequential steps summarized in Figure 1.
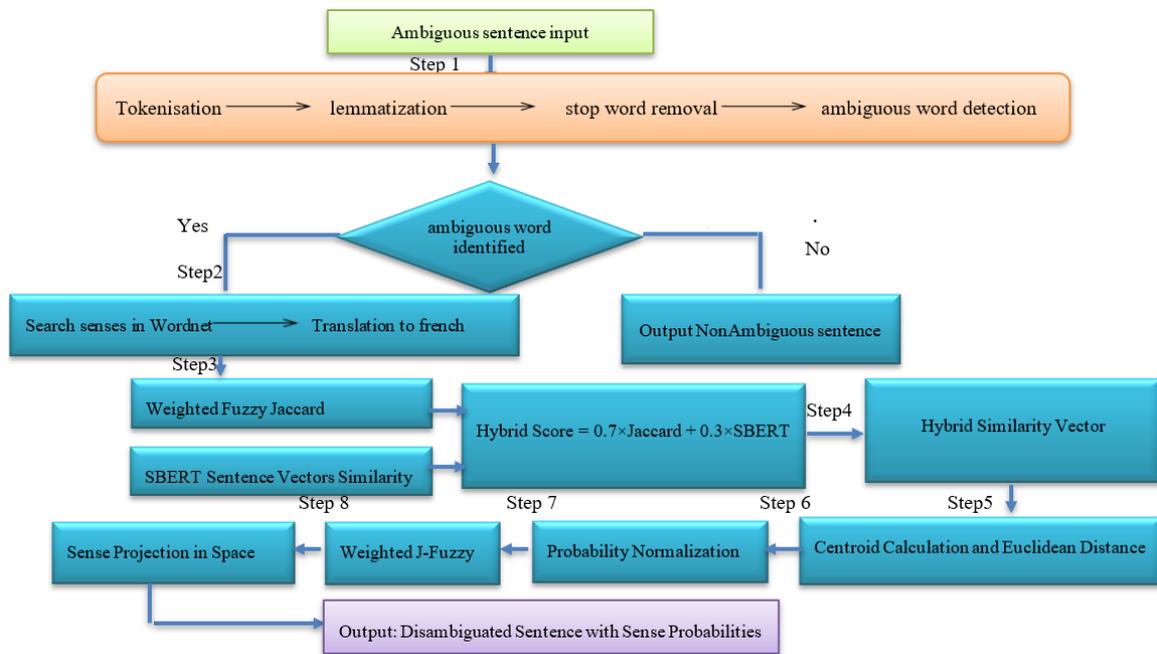

Figure 1. Workflow of the proposed hybrid approach

3.3. Step 1: token filtering and ambiguity detection

Each token in the sentence is checked against the WordNet lexical databases to detect ambiguous terms, which form the set A, while unambiguous words form the set C. For example, in the sentence "Les fils de conducteur ont coupé le courant" ("The driver's sons cut the power"), we obtain: A = {fils, conducteur} and C = {ont, coupé, courant}.

Stop words are removed, and lemmatization is applied to standardize the lexical forms. This preprocessing step ensures that the similarity calculations.

3.4. Step 2: lexical sense (Wordnet)

3.4.1. Polysemous terms

For each polysemous term, English definitions were retrieved from WordNet and translated into French, retaining their unique synset identifiers.

"fils": v_{11}: "un conducteur métallique qui transporte l'électricité sur une distance" (wire.n.02).

v_{12}: "une progéniture humaine masculine" (son.n.01).

"conducteur": v_{21}: "l'opérateur d'un véhicule à moteur" (driver.n.01).

v_{22}: "un appareil conçu pour transmettre l'électricité, la chaleur, etc." (conductor.n.04).

3.4.2. Gloss preprocessing

Glosses are preprocessed by removing French stopwords, lemmatizing nouns and verbs, and deleting occurrences of the target word to avoid bias. Example : conducteur (conductor.n.04) original gloss : "un appareil conçu pour transmettre l'électricité, la chaleur, etc." After preprocessing: "appareil concevoir transmettre électricité chaleur". This yields clean, standardized glosses for precise semantic similarity computation.

3.5. Step 3: hybrid contextual similarity mesurement

Semantic coherence between candidate senses of ambiguous words and their contextual terms is evaluated using a hybrid model integrating weighted fuzzy jaccard and SBERT embeddings.

3.5.1. Weighted fuzzy jaccard

The weighted fuzzy jaccard, adapted for French NLP, measures lexical overlap between glosses of candidate senses and context words' dominant senses. Scores range from 0 to 1, with 1 indicating perfect alignment.

$$J_{fw(A,B)} = \frac{\sum_{i \in A} \sum_{j \in B} min(w_i, w_j).s(i,j)}{\sum_{i \in A} w_i + \sum_{j \in B} w_j - \sum_{i \in A} \sum_{j \in B} min(w_i, w_j).s(i,j)} \qquad (1)$$

w_i, w_j: term weights.
A, B: preprocessed glosses of the target word sense and context word dominant sense.
$s(i, j)$: semantic similarity between terms i and j.

Output:
The weighted fuzzy Jaccard index, adapted for French NLP, measures lexical overlap between preprocessed glosses of candidate senses and the dominant context senses. Each pair is scored from 0 to 1, with 1 indicating perfect semantic alignment.
Preprocessing:
− Lemmatization ("coupés" → "couper")
− Stopword removal ("de", "le", "ont")
− Bias mitigation (exclude target word from own gloss)

3.5.2. Cosine-based semantic analysis for French using multilingual SBERT

To capture deeper semantic relationships beyond lexical overlap, multilingual SBERT embeddings compute cosine similarity between sentence vectors in high-dimensional space, effectively modeling syntactic and conceptual dependencies, with optimization for French.

$$SBERT(X, Y) = cos(\theta_{XY}) \in [0,1] \qquad (2)$$

XY: Input text representation.
θ_{XY}: angle between embedding vectors in high dimensional space.

3.5.3. Hybrid scoring formula

The contextual similarity between candidate words and their surrounding context was evaluated using a hybrid model that combines weighted fuzzy jaccard and SBERT embeddings (Table 2):

$$score_{final} = \alpha \times Jaccard + (1 - \alpha) \times SBERT \qquad (3)$$

Parameters:
− α=0.7: weight for lexical similarity using weighted fuzzy jaccard.
− 1-α=0.3: weight for semantic similarity (SBERT).

The weighting parameter (α) controls the relative contribution of lexical similarity (weighted fuzzy jaccard) and semantic similarity (SBERT). To determine the optimal value of α, a Python-based program tested values ranging from 0 to 1 in increments of 0.1, using sentence pairs that are lexically similar but semantically divergent. The program automatically selected the α value that maximized the overall similarity accuracy, resulting in an optimal α = 0.7. This configuration prioritizes lexical similarity while preserving semantic contextual nuances, achieving a balanced performance in detecting semantic divergences among lexically close sentences.

Table 2. Hybrid contextual similarity measurement

Target word	SBERT (coupé)	SBERT (courant)	Jaccard (coupé)	Jaccard (courant)	Global score
Fils	0.288	0.572	0.570	0.708	0.534
Fils	0.438	0.288	0.592	0.441	0.439
Conducteur	0.496	0.372	0.842	0.730	0.609
Conducteur	0.302	0.694	0.702	1.000	0.674

Interpretation:
- Wire (v_{11}) and conductor (v_{22}): high global score (0.563 and 0.636). Their very strong semantic SBERT and JACCARD similarity with "current". Show excellent coherence with the electrical context.
- Son (v_{12}) and driver (v_{21}): their similarities are weaker and primarily aligned with "cut", indicating a less relevant coherence for the electrical context.
- Key finding: only wire (v_{11}) and conductor (v_{22}): demonstrate optimal alignment with the sentence.

3.6. Step 4: sense weighting
For each ambiguous word sense s_i, build a hybrid similarity vector (Table 3):

$$\vec{v_i} = 0.7 \times J\,(s_i, coupe') + 0.3 \times SBERT + 0.7 \times J(s_i, courant)0.3 \times SBERT(s_i, courant) \quad (4)$$

$\vec{v_i}$: global vector representing the contextual similarity for the ambiguous word s_i.

Example:
For *"conducteur"* as *"driver"*:
$$\vec{v} = [0.7 \times 0.843 + 0.3 \times 0.497, 0.7 \times 0.729 + 0.3 \times 0.371] = [0.739, 0.622]$$

Table 3. Vector based contextual alignmment of word sense

Target word	Vector	Interpretation
Fils	[0.486, 0.666]	Strong alignment with electrical current
Fils	[0.547, 0.394]	No meaningful correlation with electrical context
Conducteur	[0.739, 0.622]	Moderate technical relevance
Conducteur	[0.583, 0.908]	Good current correlation

These vectors allow quantitative weighting of each sense based on its semantic and lexical coherence, facilitating the selection of the most contextually appropriate meaning.

3.7. Step 5: centroid analysis
Identify the sense most coherent with the overall context using a geometric center. Table 4 shows Centroid calculation: compute the arithmetic mean of the coordinates of all points (all senses combined).

$$G = \frac{1}{n}\sum_{i=1}^{n} V_i \quad (5)$$

Where:
$G = (G_x, G_y)$ is the centroid or mean center of the points.
$V_i = (x_i, y_i)$ is the vector representing the i-th point.
n is the total number of points
Example: with 4 points:

$$G = \left(\frac{0.486 + 0.547 + 0.739 + 0.583}{4}, \frac{0.666 + 0.394 + 0.622 + 0.908}{4}\right) = (0.589, 0.648)$$

Euclidean distance to the centroid:
For each sense s_i with coordinates (x_i, y_i):

$$D(s_i, G) = \sqrt{(x_i - G_x)^2 + (y_i - G_y)^2} \quad (6)$$

Example: distance for wire.n.01 (0.486, 0.666):

$$d = \sqrt{(0.486 - 0.589)^2 + (0.666 - 0.648)^2} = \sqrt{0.0109} = 0.104$$

Table 4. Euclidean distance to the centroid

Target word	sense	Distance	Interpretation
Fils	v_{11}	0.104	Very close to de centroid
Fils	v_{12}	0.257	Farthest from the centroid
Conducteur	v_{21}	0.154	Further from the centroid
Conducteur	v_{22}	0.261	Closest to the centroid

3.8. Step 6: normalization

The probability of a sense i for an ambiguous word is the ratio of the exponential of its global score to the sum of the exponentials of the global scores of all possible senses j of the same word (Table 5).

$$P(s_i) = \frac{e^{Global\ Score_i}}{\sum_j e^{Global\ Score_j}} \qquad (7)$$

Table 5. Probability

Target word	Global score	Probability
Fils	0.563	53.6%
Fils	0.417	46.4%
Conducteur	0.636	50.3%
Conducteur	0.626	49.7%

Interpretation: the SoftMax normalization then converts these global scores into probabilities, providing a relative likelihood of each sense in the given context. In the table above, the sense wire receives a probability of 53.6%, while the sense son has 46.4%. This indicates that wire is slightly more likely in the context, but both senses are still plausible. Probabilities for other ambiguous words are interpreted similarly, giving a normalized, comparative measure of sense relevance.

3.9. Step 7: computation of the global J‑fuzzy weighting score and semantic similarity measurement between lexicaly similar sentences

The global J‑fuzzy weighted score combines fuzzy lexical similarity and sense probability for WSD. Candidate definitions are scored using fuzzy Jaccard and SBERT similarity, normalized via SoftMax, and the final score is the weighted average of fuzzy Jaccard values multiplied by sense probabilities

$$J_{FuzzyWeighted} = \frac{1}{N} \sum_{i=1}^{N} P(s_i) \times Jaccard(s_i) \qquad (8)$$

Here, denotes the sense probability and represents the fuzzy lexical similarity between context and the sense definition. This measure captures semantic similarity even for lexically close sentences; for instance, "Les fils du conducteur ont coupé le Courant" does not reach full similarity because ambiguous word senses (fils: son vs fils: wires) affect the score. Thus, the global J‑fuzzy weighted score reflects both lexical overlap and sense plausibility, integrating contextual and semantic understanding, enabling distinction of lexically identical but semantically different sentences, and offering a more human-like and robust measure than purely lexical approaches.

3.10. Step 8: variant projection

To represent the potential meanings of ambiguous terms like fils and conducteur, a vector space is built with two contextual dimensions: the x-axis encodes weighted fuzzy Jaccard and SBERT similarity with coupé (electrical interruption), and the y-axis encodes the same similarity with courant (electrical flow). Interpretation:

This dual-metric approach combines fuzzy Jaccard and SBERT to position each sense in the contextual vector space. Centroid analysis shows that conducteur (électrique) and fils (électrique) are closest to the average context, confirming their technical relevance, while fils (familial) and conducteur (chauffeur) lie farther away, excluding them from the electrical domain. This vector-based analysis validates optimal disambiguation for the context. Geometric analysis of contextual word senses as shown in Figure 2.
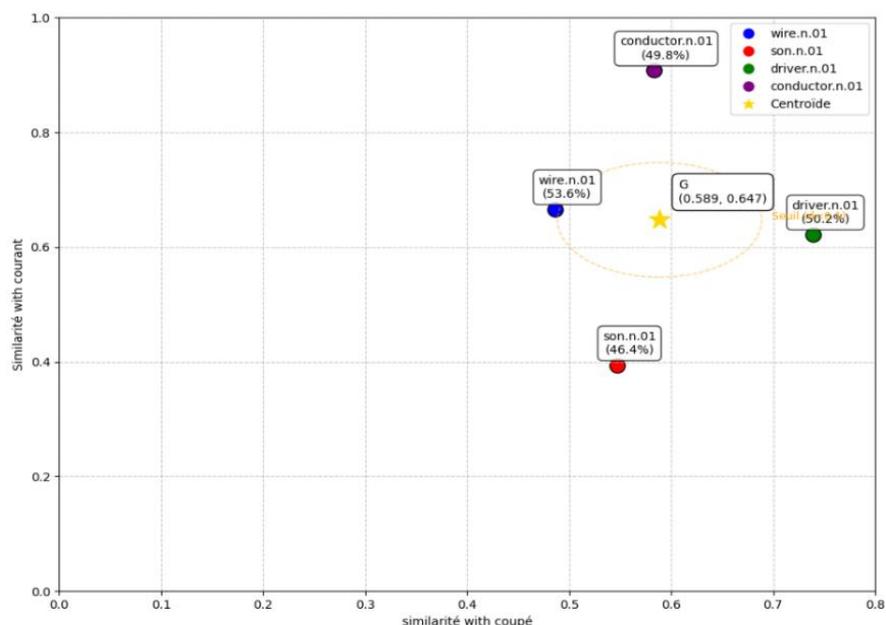
Figure 2. Geometric analysis of contextual word senses

3.11. Computational complexity and scalability

The computational complexity of the proposed model varies across its components. Tokenization and lemmatization scale linearly with sentence length, while lexicon lookup depends on the number of ambiguous tokens. The weighted fuzzy Jaccard computation scales with the size of glosses and contextual tokens, and SBERT inference depends on both sentence length and embedding dimensionality. Centroid and distance calculations scale with the number of candidate senses, and the SoftMax layer with the number of sense probabilities to compute.

From a practical standpoint, efficiency can be improved by caching embeddings and intermediate results, performing batched SBERT inference on GPU, and parallelizing computations across tokens or sentences. These optimizations ensure scalability for larger datasets and longer textual inputs.

3.12. Interpretability

Our approach provides interpretability at three complementary levels:
− Lexical level: the weighted fuzzy Jaccard identifies key tokens in WordNet definitions, revealing their relative importance in similarity calculation.
− Vector and distance level: sentences are represented as hybrid vectors combining lexical and semantic features. The centroid serves as a semantic reference point, where shorter Euclidean distances indicate higher similarity and longer distances show semantic divergence.
− Probabilistic level: the SoftMax function converts raw similarity scores into probability distributions, enabling clear comparison between synsets and providing a probabilistic basis for component fusion.

4. EXPERIMENTAL VALIDATION
4.1. Human evaluation framework

A rigorous evaluation protocol was implemented using a 0-100 scale with native French speakers, including linguistics and NLP specialists. The methodology demonstrated robust reliability with Fleiss' Kappa of 0.48 (moderate agreement) and test-retest correlation of 0.82, supported by comprehensive quality control measures to ensure data validity.

4.2. Comparative systme specification

The experimental evaluation incorporated comprehensive state-of-the-art systems including GPT-4 (gpt-4-0613), Gemini Pro (gemini-1.5-pro), Perplexity AI (2024-08 API), DeepSeek Chat (deepseek-llm-67b-chat), along with research systems SimCSE-fr and T5-Encoder. All systems underwent identical preprocessing pipelines with unified 768-dimensional embeddings, using cosine similarity with L2 normalization.

4.3. Experimental results

A comprehensive analysis of 33 sentences compared human evaluations with AI-generated similarity scores and our hybrid approach, as detailed in Table 6.

Table 6. Evaluation human judgments vs artificial intelligence (AI) systems vs hybrid approach

Sentences	Approach	Human	AI	Human-AI	Human-approach
Les fils du conducteur ont coupé le courant	0.72	0.32	1.00	0.79	0.43
Le médecin a demandé une radio du bras	0.73	1.00	1.00	0.00	0.61
La carte du roi est perdue	0.37	0.04	1.00	0.62	0.72
La police est vue avec une barre	0.24	0.61	1.00	0.39	0.62
Le responsable de notre poste a visité le parc	0.56	0.81	1.00	0.18	0.82
Il faut ajouter plusieurs serveurs pour renforcer notre système	0.73	0.56	1.00	0.44	0.59
Le professeur a cassé la règle en classe	0.79	0.00	1.00	1.00	0.00
Après le vol de la nuit, les balles sont lancées au ciel	0.98	0.64	1.00	0.37	0.23
Les feuilles sont sauvegardées sur la table	0.37	0.86	1.00	0.12	0.83
La table est très utile pour le travail au bureau	0.87	0.77	1.00	0.23	0.34
Le tour du roi est dangereux	0.12	0.49	1.00	0.51	0.60
La fenêtre est restée ouverte pendant qu'il marche	0.86	0.76	1.00	0.24	0.33
Les puces sont visibles sur la carte	0.00	0.28	1.00	0.83	0.83
On a découvert la valeur de notre tour au cours de l'histoire	0.59	0.57	1.00	0.38	0.59
La somme des charges est lourde	0.17	0.32	1.00	0.68	0.41
La roue a roulé près du train	0.97	0.65	1.00	0.25	0.05
Sa voix compte dans la décision finale	0.83	0.56	1.00	0.45	0.25
Il a posé une couche sur la table	0.65	0.40	1.00	0.61	0.31
Le plan du spectacle a été imprimé sur la table	0.74	0.58	1.00	0.43	0.24
Il a acheté une livre au marché avec ses fils	0.40	0.79	1.00	0.21	0.49
La mine est usée après beaucoup d'usage	0.43	0.66	1.00	0.33	0.44
La pression sur le cours est inquiétante	0.17	0.56	1.00	0.45	0.92
Les enfants portent une orange à la maitresse près du parc	0.42	0.75	1.00	0.33	0.32
Elle a écrit la lettre en capitale ce matin	0.49	0.34	1.00	0.67	0.31
Il a perdu la pile au cours du match	0.45	0.31	1.00	0.70	1.00
La manche a pris près du feu	1.00	0.69	1.00	0.31	0.10
La branche qui mène au droit est bien définie	0.64	0.21	1.00	0.69	0.39
Le voile est tombé au moment du spectacle	0.43	0.41	1.00	0.60	0.21
Java possède plusieurs ports actifs à gérer	0.42	0.57	1.00	0.44	0.43
Durant la saison de récolte, la pêche a cassé la branche	0.44	0.58	1.00	0.25	0.41

Statistical analysis shows that the difference between AI systems and human evaluations is highly significant ($p = 0.000059 < 0.001$). Thus, the null hypothesis (H_0) is rejected, indicating that current systems significantly diverge from human judgments, while the proposed hybrid approach achieves better convergence. The effect size (Cohen's $d = -0.80$) confirms a strong effect, demonstrating that the improvement is both statistically and practically meaningful.

Confidence intervals further support the robustness of these findings:
− MAE (Hybrid approach): 40.02 [39.05 – 40.99].
− MAE (Reference AI): 45.32 [43.39 – 47.26].

The average reduction of 5.31 points corresponds to a relative improvement of 11.7%. Since the confidence interval [-7.64, -2.97] excludes zero, the observed effect is robust and reliable. Moreover, the proposed approach achieves the lowest MAE and root mean squared error (RMSE), the highest semantic accuracy (0.875), and the best mean rank (1.8), confirming its stability and closeness to human reasoning. Table 7 compares the hybrid approach with SimCSE and T5 across multiple metrics, assessing semantic alignment, error minimization, and correlation with human judgments:

While T5 achieves higher correlation coefficients due to its large-scale pretraining, the hybrid approach outperforms both models in MAE, RMSE, and semantic accuracy, confirming its greater stability and interpretability. The lower error margins indicate that it not only captures general semantic trends but also fine-grained distinctions in meaning.

Table 7. Comparative performance-hybrid approach vs.modern methods

Metric	Hybrid approach	SimCSE	T5	Best
Spearman correlation	0.762	0.714	0.905	T5
Pearson correlation	0.841	0.701	0.926	T5
MAE	0.133	0.309	0.376	Hybrid
RMSE	0.211	0.408	0.465	Hybrid
Semantic accuracy	0.875	0.500	0.500	Hybrid
Mean rank	1.8	2.8	2.4	Hybrid

5. DISCUSSION

Our hybrid approach demonstrates superior semantic coherence with human judgments through strong correlations (Spearman: 0.762, Pearson: 0.841) and significantly reduced errors (MAE: 0.133 vs 0.309-0.376 for baselines), achieved via adaptive integration of symbolic and neural components that dynamically responds to ambiguity patterns. While T5 achieves higher correlations (0.905 Spearman), it systematically overestimates lexical similarity, whereas our method maintains both contextual sensitivity and crucial interpretability, despite facing limitations in cross-linguistic adaptation and handling complex pragmatic ambiguities. These findings challenge conventional scaling paradigms by demonstrating that sophisticated architectural design can achieve superior performance compared to simply increasing model parameters, thereby enabling more resource-efficient deployment.

6. CONCLUSION

This study presents a hybrid approach for French semantic similarity and WSD, combining fuzzy Jaccard lexical matching with SBERT embeddings. Experimental validation on 33 ambiguous sentences, evaluated by native speakers, demonstrated that the hybrid model significantly outperforms existing AI systems, achieving an 11.7% reduction in MAE compared to ChatGPT and Gemini (t = -4.62, p = 5.92×10^{-5}). The approach produces interpretable, context-sensitive results, particularly for lexically identical but semantically divergent sentences, demonstrating strong alignment with human judgments and transparent decision-making.

Our approach changes the state of the art by introducing dynamic integration between symbolic and neural methods, whereas previous work favored static combinations or purely neural architectures. This innovation enables contextual ambiguity resolution that surpasses the limitations of conventional approaches, while maintaining the interpretability often sacrificed in large-scale models.

The hybrid framework contributes theoretically by bridging symbolic linguistic knowledge with neural contextual modeling, and practically by improving NLP tasks such as machine translation, question answering, and semantic search with accurate and explainable sentence-level representations. Limitations include the restricted evaluation set and focus on French, while future perspectives involve multilingual adaptation, scaling to larger benchmarks, and integration of explainable modules to enhance interpretability and user trust.

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 El Janati	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
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 they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

INFORMED CONSENT

Informed consent was obtained from all participants included in this study. Participants were informed about the purpose of the study, and their participation was voluntary. All data collected were anonymized to ensure privacy and confidentiality.

ETHICAL APPROVAL

Ethical approval was not required for this study as it involved voluntary human judgments on de-identified textual data. Informed consent was obtained from all participants included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [initials], upon reasonable request.

REFERENCES

[1]   H. Le *et al.*, "FlauBERT: unsupervised language model pre-training for French," *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, pp. 2479–2490, 2020.
[2]   A. Enaanai, A. Doukkali, and B. Regragui, "Noise abatement in the Arabic IRS results by applying a sentences morphosemantic filter (gene filter method)," *Journal of Theoretical and Applied Information Technology*, vol. 42, no. 1, pp. 113–121, 2012.
[3]   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, Dec. 2017, doi: 10.1162/tacl_a_00051.
[4]   P. J. L. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, "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, 2023, doi: 10.5555/3455716.3455856.
[5]   K. Li, O. Patel, F. Viégas, H. Pfister, and M. Wattenberg, "Inference-time intervention: eliciting truthful answers from a language model," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
[6]   L. Martin *et al.*, "CamemBERT: a tasty french language model," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7203–7219, doi: 10.18653/v1/2020.acl-main.645.
[7]   S. Arora, Y. Li, Y. Liang, T. Ma, and A. Risteski, "A latent variable model approach to PMI-based word embeddings," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 385–399, 2016, doi: 10.1162/tacl_a_00106.
[8]   J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
[9]   Y. Guo, "BERTweetFR: a pre-trained model for french tweets," in *Proc. Workshop Noisy User-generated Text*, 2021, pp. 445–450.
[10] C. Delestre and A. Amar, "DistilCamemBERT: a distilled version of CamemBERT," *arXiv preprint arXiv:2205.11111*, 2022.
[11] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTaV3: improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing," *International Conference on Learning Representations (ICLR-2021)*, pp. 1–16, 2021.
[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] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT sentence embedding," *arXiv.2007.01852*, 2020, [Online]. Available: https://arxiv.org/abs/2007.01852.
[14] T. Gao, X. Yao, and D. Chen, "SimCSE: simple contrastive learning of sentence embeddings," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 6894–6910, doi: 10.18653/v1/2021.emnlp-main.552.
[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 2013.
[16] C. Su, H. Zheng, D. Peng, and X. Wang, "DiCA: disambiguated contrastive alignment for cross-modal retrieval with partial labels," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 19, pp. 20610–20618, Apr. 2025, doi: 10.1609/aaai.v39i19.34271.
[17] M. Artetxe, S. Ruder, and D. Yogatama, "On the cross-lingual transferability of monolingual representations," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4623–4637, doi: 10.18653/v1/2020.acl-main.421.
[18] L. Gan, Z. Teng, Y. Zhang, L. Zhu, F. Wu, and Y. Yang, "SemGloVe: semantic co-occurrences for GloVe from BERT," *arXiv.2007.01983*, 2020, [Online]. Available: https://arxiv.org/abs/2012.15197.
[19] 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.
[20] 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*, pp. 11849–11858, 2024, doi: 10.18653/v1/2024.lrec-main.1041.
[21] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.
[22] M. Zhang, M. Mosbach, D. I. Adelani, M. A. Hedderich, and D. Klakow, "MCSE: multilingual contrastive sentence embeddings," *arXiv.2303.01815*, 2023.
[23] A. Vaswani, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, pp. 5998–6008, 2017, doi: 10.5555/3295222.3295349

BIOGRAPHIES OF AUTHORS

Btissam El Janati holds a bachelor's degree in Physics, a Master's degree in Embedded Electrical Systems, and a Ph.D. in Computer Science specializing in NLP. Her research focuses on automatic keyword extraction, lexical disambiguation, and multimodal approaches for French language processing. She has published 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 the cutting-edge fields of information retrieval systems, big data, and artificial intelligence. Born in Fez in 1981, he pursued his higher education at ENSIAS in Rabat, where he earned his Ph.D. in Computer Science in 2014. As a dedicated teacher-researcher, he combines educating future generations with advanced research, focusing on the scientific and technological challenges of data management, search algorithms, and AI applications. His Habilitation to Lead Research (HDR) qualifies him to supervise doctoral theses, solidifying his role as a key contributor to the academic and research community. He can be contacted at email: enaanai@uae.ac.ma.

Fadoua Ghanimi She is a professor at Ibn Tofail University, Morocco, specializing in computer science and artificial intelligence applied to the IoT and smart agriculture. Her research focuses on smart 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.