Correcting Sense Annotations using Wordnets and Translations

Arnob Mallik and Grzegorz Kondrak

Global WordNet Conference
January, 2023

Preliminaries : WSD

Word Sense Disambiguation (WSD) : The task of identifying the correct sense of a word in context, given a predefined sense inventory (Navigli, 2009).

Example:

The bat is feeding on fruit.

Nocturnal mammal

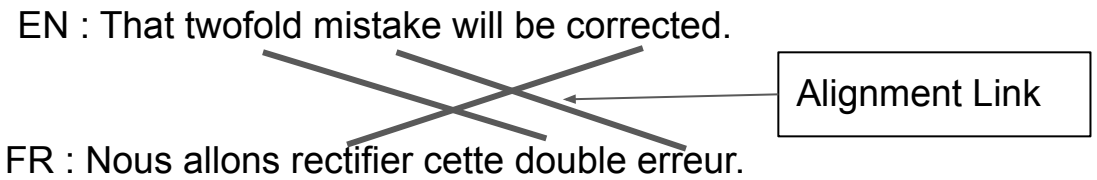
He hit the ball with the bat.

A club used for hitting a ball in various games.

Preliminaries : Bitexts and Word Alignment

Bitexts : EuroParl, OpenSubtitles → source of translations.

Word alignment tools are employed on **bitexts** to extract word-level translations.



BabAlign (Luan et al. 2020) : High precision alignment algorithm. Improves the output of a base aligner (e.g. **FastAlign**) by utilizing BabelNet information.

Approach

We exploit two different properties of translations :

- Property 1 (Equivalence) : A word and its translation, in most cases, should represent the **same concept**.
  - Improve sense annotations.
  - Build WSD pipelines.
- Property 2 (Generalization) : Words, in some cases, can be translated into **more general concepts**.
  - Cross-lingual lexical entailment.
    - EN : You gave me the bottle.
    - IT : Mi hai dato il contenitore.

Prior Work : Sense Annotations from Translations

- ➔ **Resnik and Yarowski (1997)**: Translation distinctions may correlate to sense distinctions. Example : *duty*^{EN} translated into *devoir*^{FR} (obligation) or *droit*^{FR} (tax).
- ➔ **Chan and Ng (2005)**: semi-automatically disambiguated English nouns using Chinese translations retrieved from an English-Chinese parallel corpus.
- ➔ **Deli Bovi et. al. (2017)**: Proposed an automatic approach of jointly disambiguating multiple languages of a parallel corpus.
- ➔ **Luan et. al. (2020)**: Proposed approaches of improving the output of a base WSD system by leveraging translations.

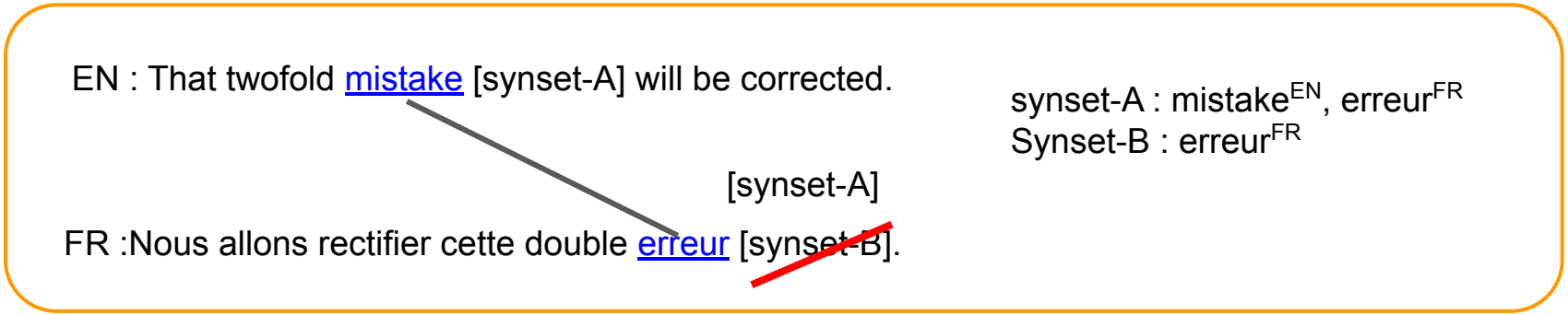
Our Approach

We propose two algorithms to make **selective** corrections on an automatically sense-annotated bitext:

- ➢ **MultiWordNet (MWN) Algorithm** : Operates on individual alignment links.
- ➢ **Bipartite (BP) Algorithm** : Considers all alignment links in corpus and makes corrections based on most frequent links.

MultiWordNet (MWN) Algorithm

- **Common multi-synset** : A multilingual synset that **contains both words** in an alignment link.
- MWN Algorithm makes corrections where an alignment link involves only **one** common multi-synset.



MultiWordNet Algorithm

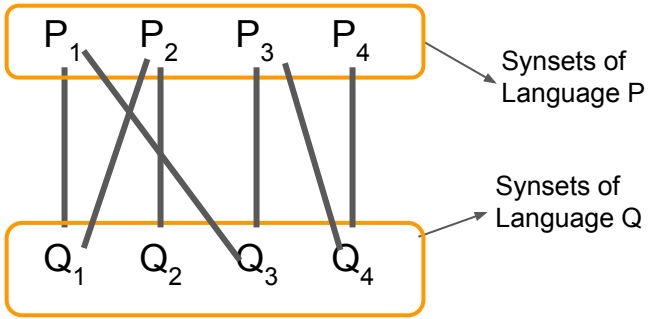
Algorithm 1 MULTIWORDNET Algorithm

Input : Aligned Sense Pair (s,t).
$w(s) \leftarrow$ Word of which s is a sense
$M(s) \leftarrow$ Multi-Synset that contains sense s
$M(w) \leftarrow$ Set of multi-synsets that contain word w.

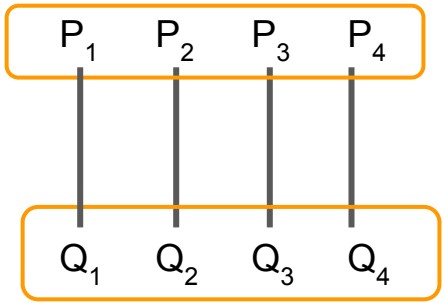
1: $C \leftarrow M(w(s)) \cap M(w(t))$
2: **if** $M(s) \neq M(t)$ **then**
3: **if** $M(s) \in C$ and $M(t) \notin C$ **then**
4: $t \leftarrow (w(t), M(s))$
5: **end if**
6: **if** $M(t) \in C$ and $M(s) \notin C$ **then**
7: $s \leftarrow (w(s), M(t))$
8: **end if**
9: **end if**

Bipartite (BP) Algorithm

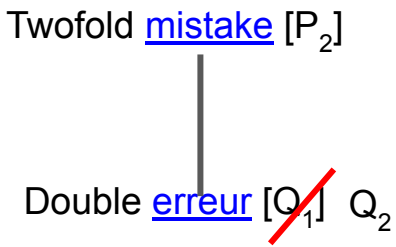
- Assumes a bipartite graph of synsets → **vertices** can be divided into two disjoint sets, each containing synsets of a particular language. **Edges** represent alignment links.
- Creates a 1 to 1 mapping of synsets based on **most frequent alignment links**. **Objective** : create a mapping of similar concepts across languages.
- Makes corrections based on the mapping.



Initial Bipartite Graph



Mapping based on most frequent links



Annotation Correction (Base Language P)

Bipartite Algorithm

Algorithm 2 BIPARTITE Algorithm

Input: Alignment links involving synset pairs $(p_1,q_1),(p_2,q_2),\ldots,(p_m,q_m)$, where $p_i \in$ *Language P* and $q_i \in$ *Language Q*
Input : Frequency Threshold α

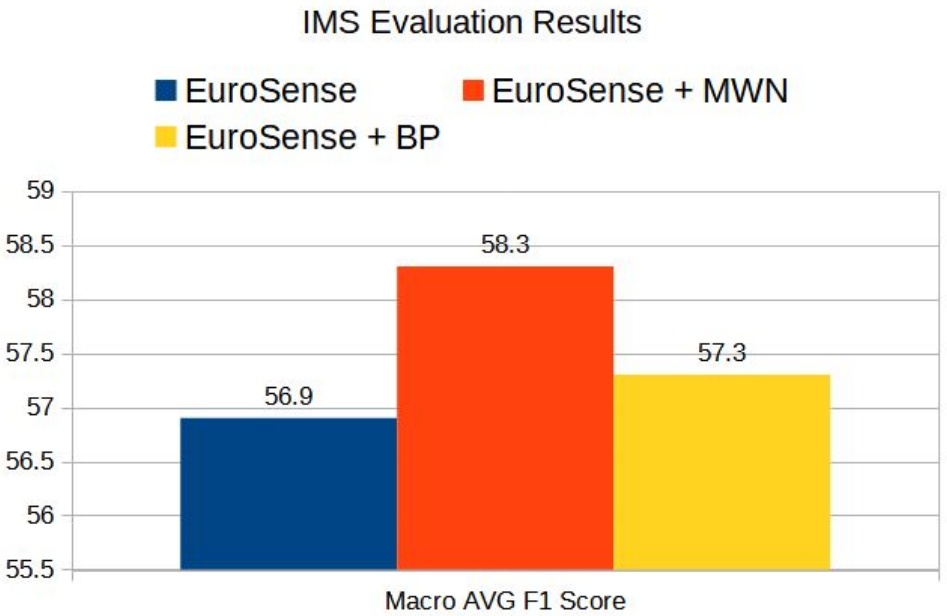
```
candidate_edges_P ← ∅, candidate_edges_Q ← ∅
Initialize Graph G (E), where Edges E ← ∅

for each language L in (P,Q) do
    for each synset x in Language L do
        n ← total alignment links involving x
        for each synset y aligned to x do
            a ← total alignment links involving (x,y)
            if a ÷ n > α then
                candidate_edges_L ← candidate_edges_L ∪ (x, y)
            end if
        end for
    end for
end for

E ← candidate_edges_P ∩ candidate_edges_Q

for  each edge (p,q) in E do
    for each synset q_i aligned to p do
        w_q ← associated word of Language Q
        if q_i ≠ q and w_q ∈ q then
            CORRECT: Alignment Link (p, q_i) ⟹ (p, q)
        end if
    end for
end for
```

Experimental Setup

- Sense-annotated corpora : **EuroSense**, based on EuroParl corpus. Annotated with **BabelNet** synsets. We extract 4 bilingual slices : EN - IT, EN - FR, EN - ES, EN - DE.
- Word Alignment : **BabAlign** (Luan et. al. 2020). After this step, we get **aligned sense pairs**.
- Filtering sense pairs :
  - Entailment : Filter out pairs if one of the synset is a hypernym of the other. Using BabelNet hypernymy links.
  - Non-literal translation : Filter out a pair if the involved words do not have a synset in common.

Extrinsic Evaluation : WSD

- Applied algorithms separately to EuroSense.
- Provided the original and corrected corpus as training data for IMS (Zhong et. al. 2010), a supervised WSD system.
- Tested on SemEval - 13 and SemEval - 15 WSD test sets.



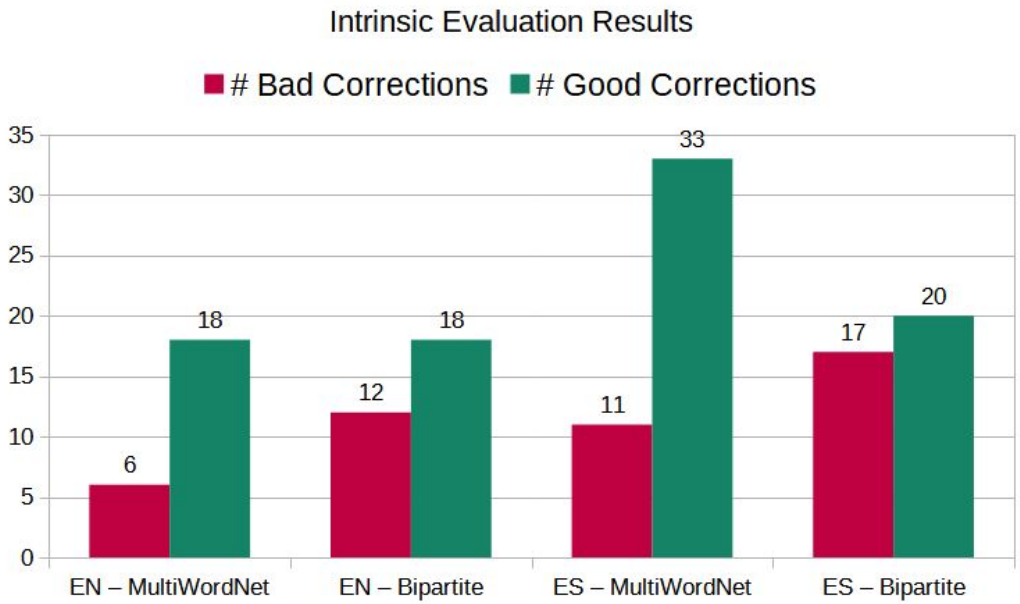
Annotation Correction Results

Training Set	Test Set							
	SemEval 2015			SemEval 2013				
	EN	IT	ES	EN	IT	FR	DE	ES
EuroSense	64.3	56.3	54.3	65.3	56.5	45.4	58.8	53.9
ES + MultiWordNet	65.1	57.1	55.3	65.5	58.3	48.0	60.0	56.7
ES + Bipartite	64.5	57.2	55.3	65.4	56.7	45.9	59.1	54.1

Table 3.2: WSD F-score of IMS trained on different corpora

Intrinsic Evaluation : Manual Annotations

- Done for **EN** and **ES**. Annotators were native speakers.
- Annotators examined the sentence containing the focus word.
- They had to pick between 3 options : the original annotation, the corrected annotation or neither.
- 100 instances per language, 50 for each algorithm.



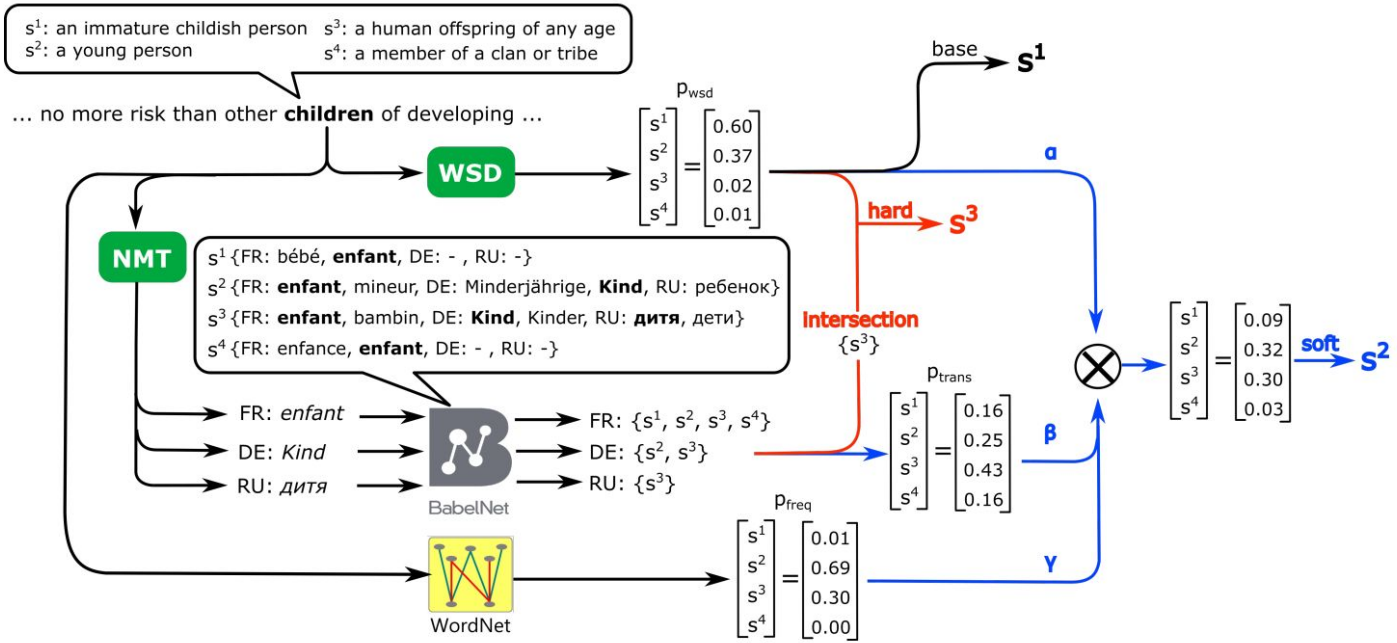
Summary

- ➢ Results constitute a strong proof-of-concept that translations can be leveraged to make effective annotation corrections.
- ➢ Error Causes:
  - **Incompleteness of BabelNet** : Some BabelNet synsets do not contain all possible lexicalizations of the concept it represents. For example, bn:00109131a contains futuro^{ES}, but not its English translation future^{EN}. Affects MWN algorithm.
  - **Significant amount of noise in EuroSense** : ~44.5 % concepts represented in English, does not exist on the German side. Affects the BP algorithm.

Thank you!
Questions?

Reference Slides

SoftConstraint (Luan et. al. 2020)



$$\tilde{p}(s) = p_{wsd}(s)^{\alpha} \cdot p_{trans}(s)^{\beta} \cdot p_{freq}(s)^{\gamma}$$

Unsupervised Corpus Labelling Using Translations

Background : WSD Systems

- Supervised Systems :
  - Examples : IMS, EWISER, GlossBert.
  - Rely on sense-annotated training data. **SemCor** (Miller et. al. 1993) -> manually annotated corpus.
  - Typically outperform knowledge-based systems.
  - Severe lack of high-quality training data, for languages other than English. Known as **knowledge acquisition bottleneck** (Pasini, 2020).
- Knowledge-Based Systems:
  - Examples : UKB, Babelfy, SensEmBERT.
  - Rely on a Lexical Knowledge Base (LKB), such as WordNet or BabelNet.
  - Lower accuracy, but higher scalability.

Prior Work : Automatic Corpus Labelling Approaches

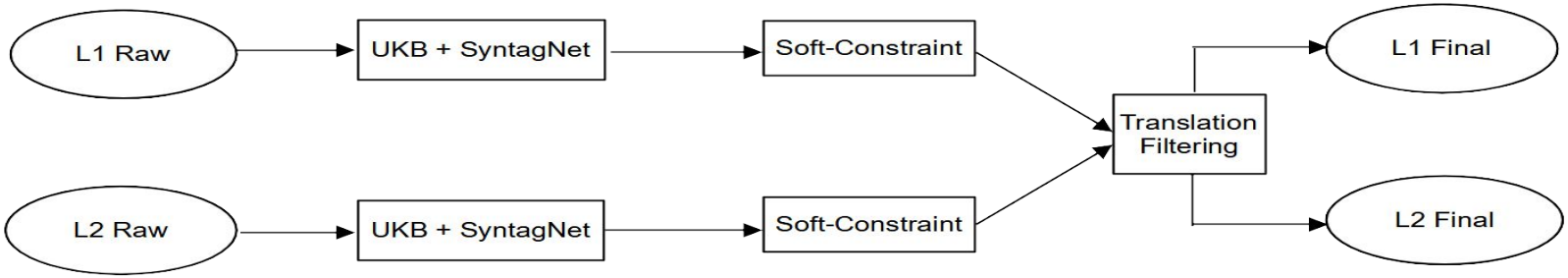
- Semi-supervised : **MuLaN (Barba et al 2020)** -> Propagates annotations from SemCor and WNG to similar contexts in Wikipedia, using contextual embeddings.
- Unsupervised:
  - **EuroSense** : Jointly disambiguates a parallel corpora using Babelfy, refines initial annotations using vector representations.
  - **Train-O-Matic** (Pasini and Navigli, 2017) : Annotates Wikipedia in multiple languages by applying PPR algorithm to BabelNet.
  - **OneSec** (Scarlini et. al. 2019): Combines representations of Wikipedia categories and BabelNet synsets to produce multilingual annotated data.

Our Approaches

We propose **fully-unsupervised** pipelines for automatically generating sense-tagged corpora :

- ➢ **LabelSync** : Language-independent approach that produces sense-annotated corpora in two languages at once by applying a KB WSD system on each side of an input bitext.
- ➢ **LabelGen** : Leverages advancements in English WSD to improve multilingual annotations.

LabelSync → Overview



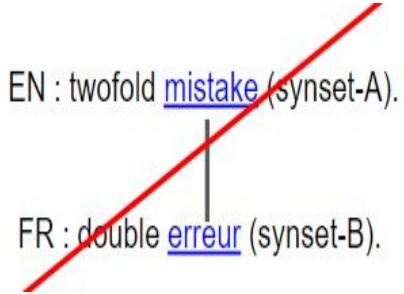
Initial WSD :

- Variant of UKB, enriched with SyntagNet (Maru et. al. 2019) on both sides of the bitext.
- Assigns a score to each sense.

Re-ranking senses:

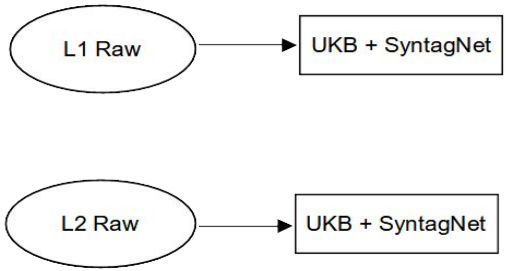
- SoftConstraint (Luan et. al. 2020) → operates on output of base WSD system.
- Depends on word-level translations → retrieved using BabAlign.

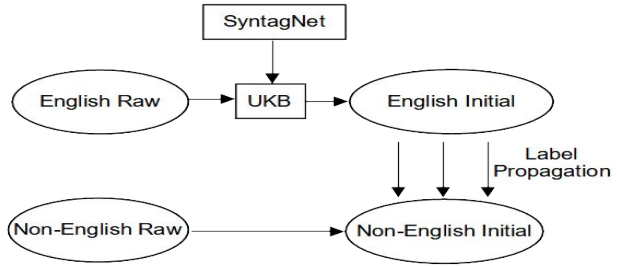
Translation Filtering :



LabelGen - Motivation

- Modified version of LabelSync, to improve multilingual annotations, by leveraging English resources. **One side of the input bitext must be English**.
- For English, UKB runs entirely using **WordNet** → reliable. For non-English, synset lexicalizations retrieved from **BabelNet** → "sub-optimal" (Scozzafava et. al. 2020)
- **LabelGen avoids running UKB on the non-English side of the bitext**.

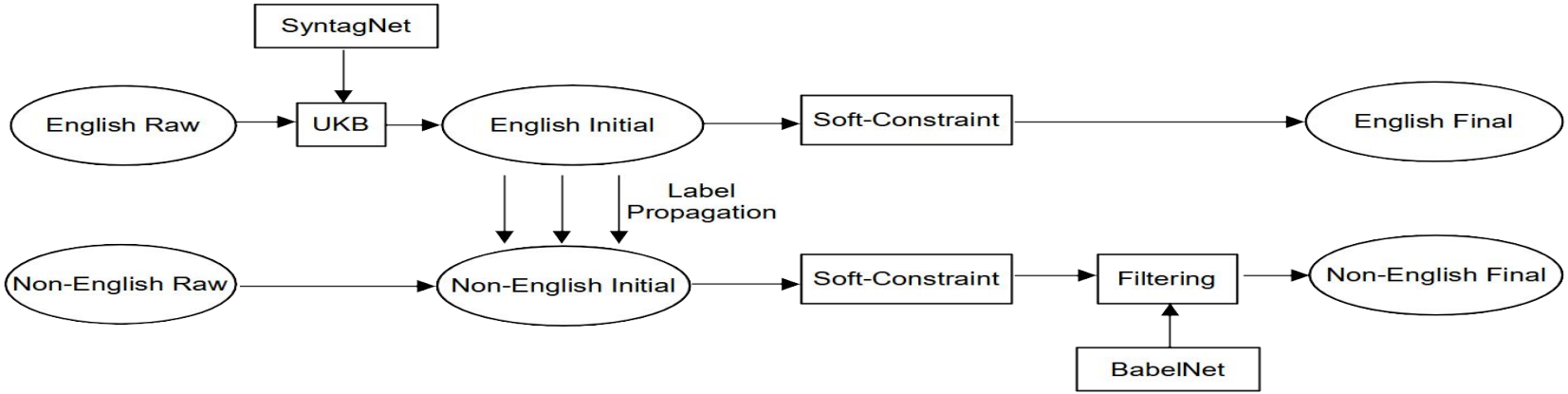




[LabelSync runs UKB on both sides]

[LabelGen runs UKB on the English side]

LabelGen - Overview



English Tagging and Label Propagation:

- UKB + SyntagNet on English
- Translations are annotated with same sense as its source.
- Scores are also propagated.

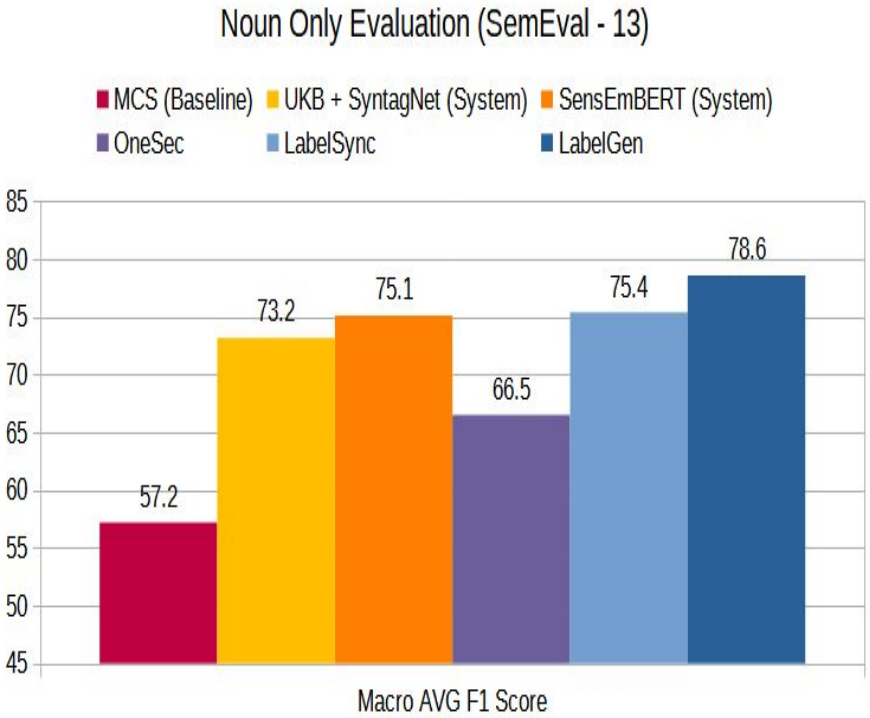
Re-ranking and Filtering:

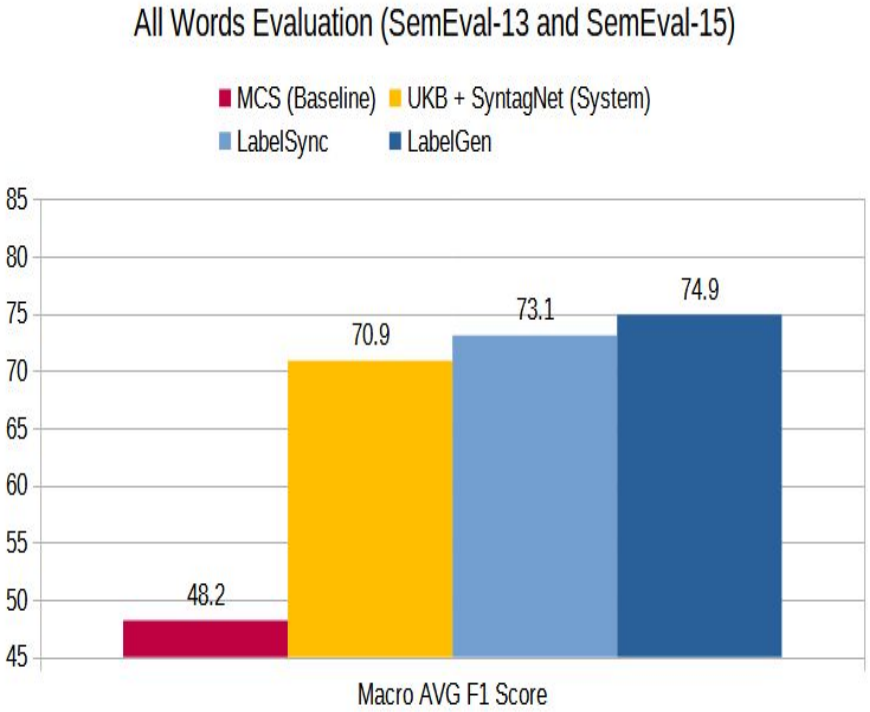
- Re-rank the senses using SoftConstraint.
- Filter out invalid annotation using BabelNet. Occurs due to alignment errors, non-literal translations.

Experiments

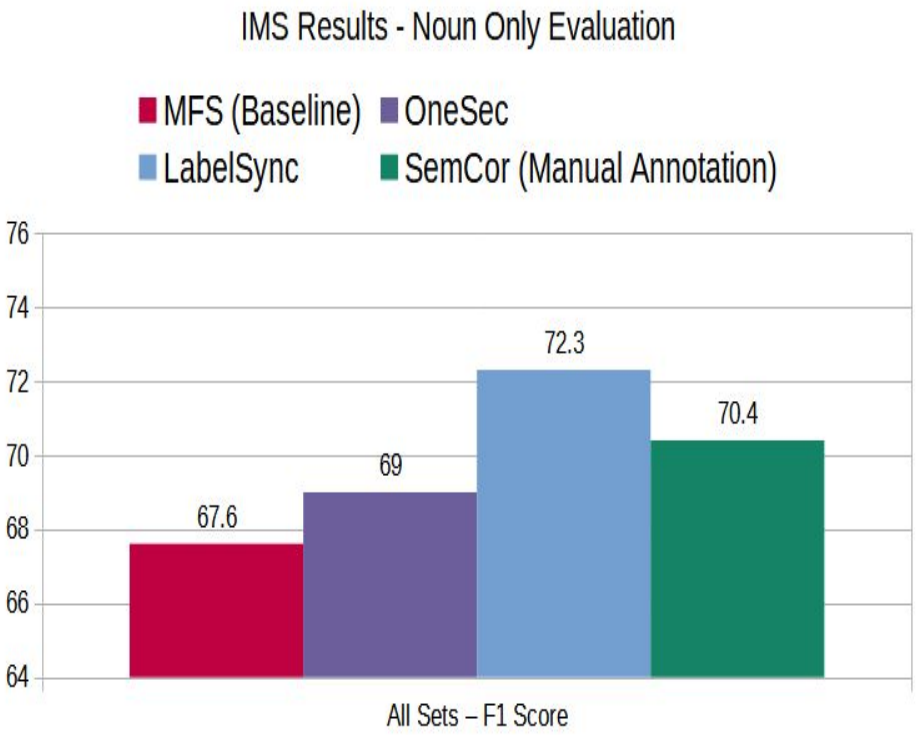
- **Input Corpus**: A subset of EuroSense, containing EN, IT, FR, DE, ES sentences → 5 language parallel corpus. We discard their annotations.
- **Extrinsic Evaluation**: We provide the annotations produced by LabelSync and LabelGen as training data for reference WSD systems :
  - mBERT (Barba et. al. 2020): transformer based system.
  - IMS (Zhong et. al. 2010): SVM based system.
- **Test Bed**: Standard benchmark datasets for multilingual and English WSD.
- **Primary Competitor** : OneSec (Scarlini et. al. 2019) → sense-annotated corpus built in an unsupervised manner.

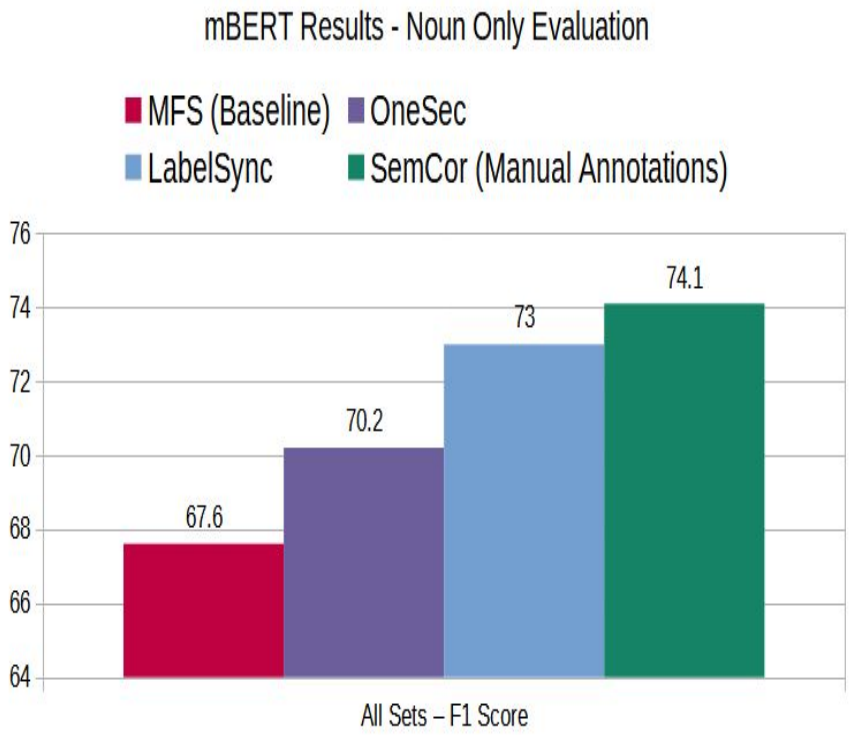
Multilingual Results (IT, ES, FR, DE) → mBERT



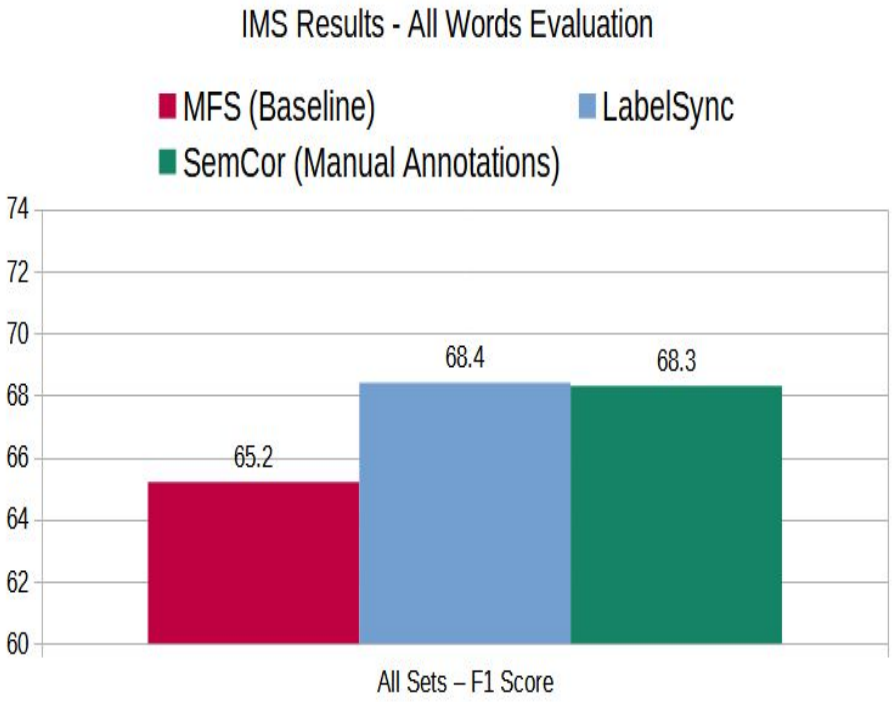


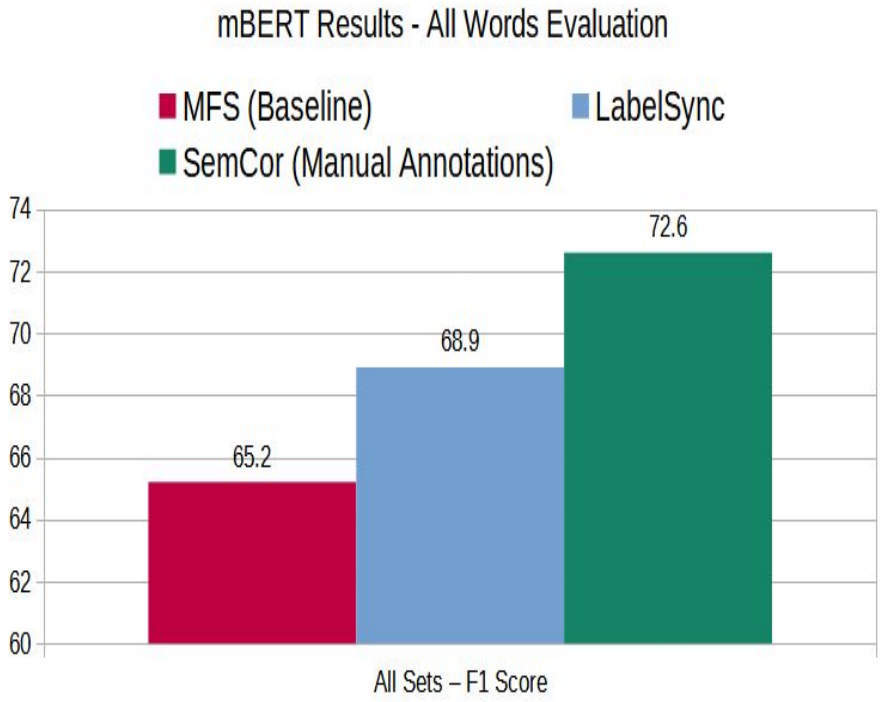
English Results → Noun Only Evaluation



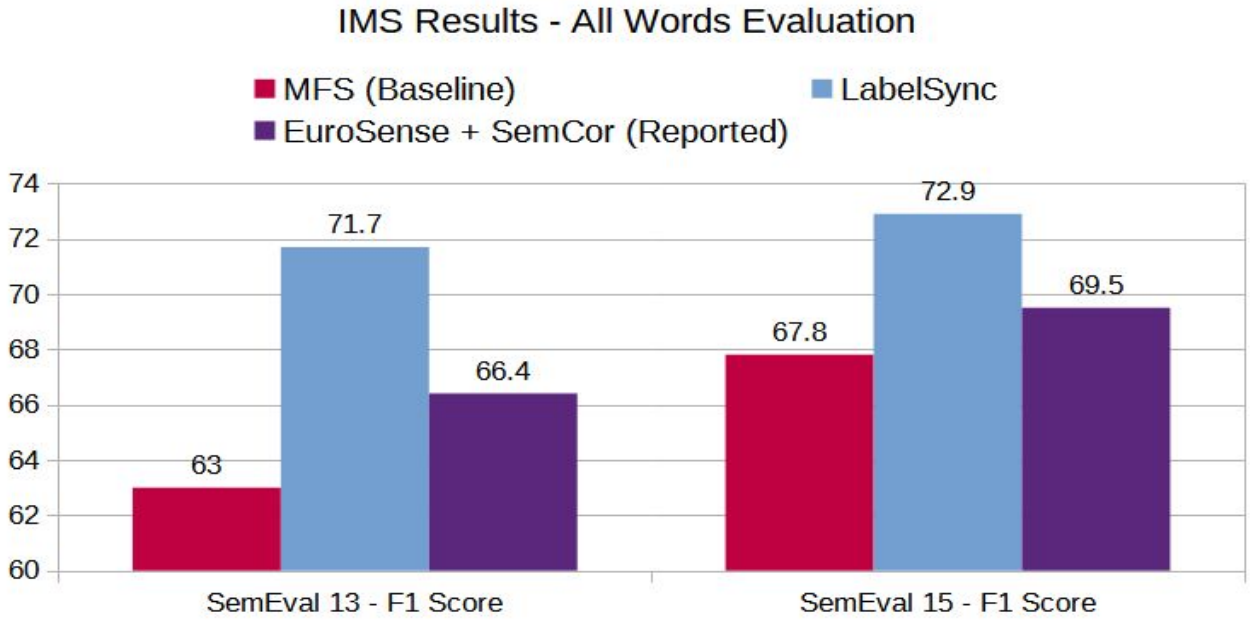


English Results → All words evaluation





English Results → Comparison with EuroSense



Summary

- Our proposed methods for automatic sense-tagging can produce annotated data for arbitrary languages and domains → a step towards mitigating **knowledge acquisition bottleneck**.
- State-of-the-art results for unsupervised multilingual WSD.
- Our annotations approach the quality of manual annotations.

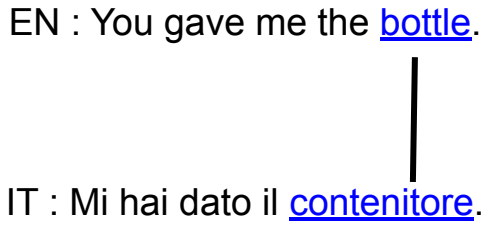
Using Translations to Predict Cross-Lingual Lexical Entailment

Task Description

- **Cross-lingual binary lexical entailment** (SemEval 2020 Task 2, Gravas et. al. 2020)
- “Detect whether the meaning of one word can be inferred from the meaning of a word in another language” → Vyas et. al. 2016
- (Jug^{EN}, $Contenitore^{IT}$) —predict→ positive / negative.

Our Objective

➢ Provide evidence for the hypothesis that translations are useful in predicting cross-lingual lexical entailment.

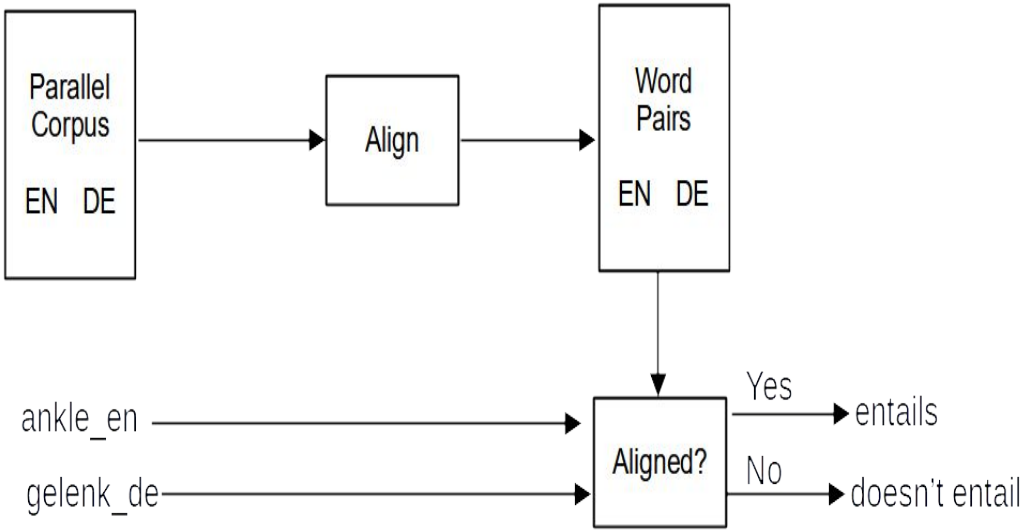


Baseline : Bitext Method

- ➢ Retrieve translation pairs from a bitext using word alignment.
- ➢ At test time, check if the word pair is in the list of translation pairs.
- ➢ If so, there is an entailment relation between the words.

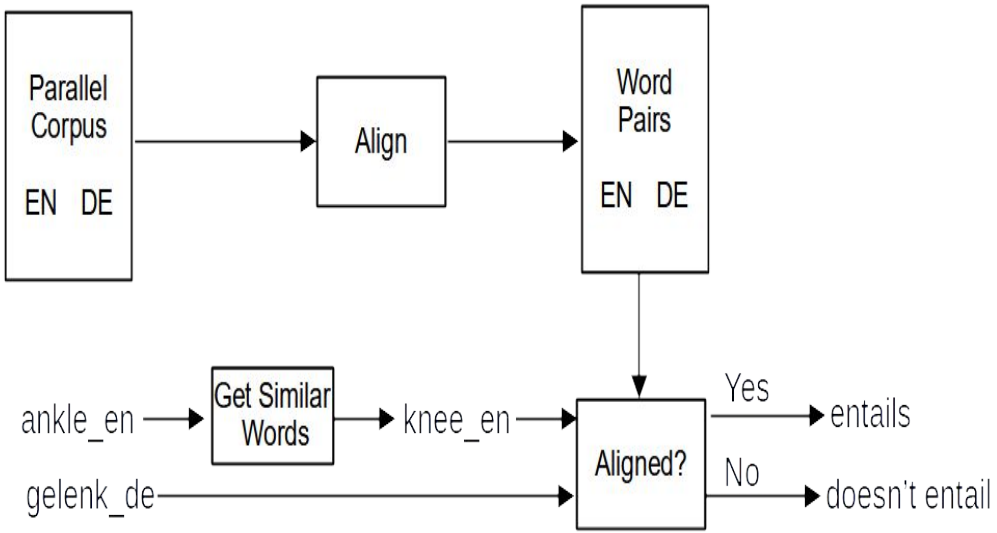
Problem :

- ➢ Constrained by the coverage of translation pairs retrieved from the bitext.



Semantic Expansion : Vectors Method

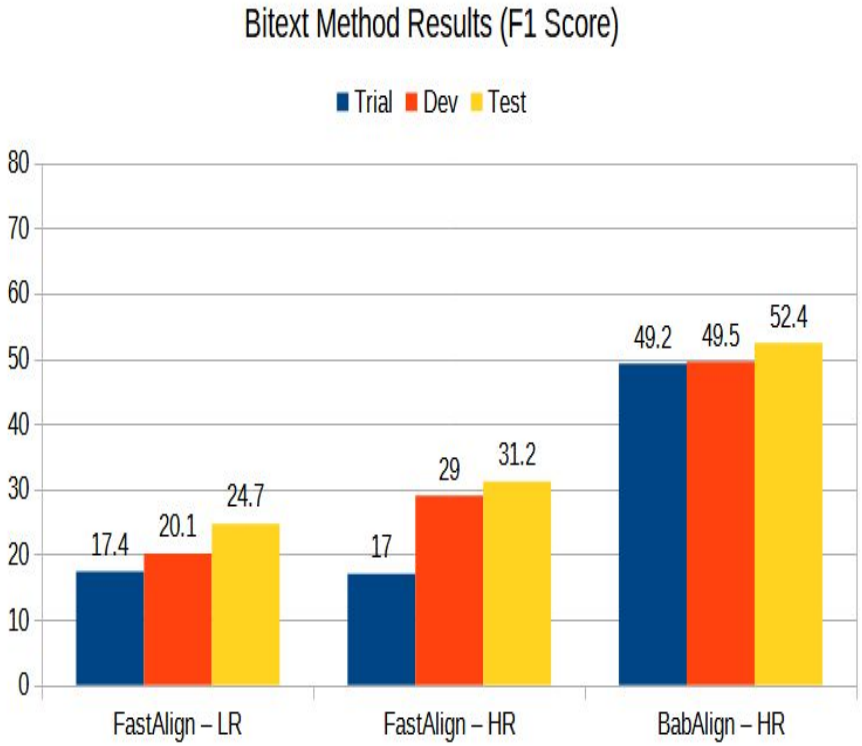
- Objective : Relax the dependency on the bitext.
- We compute cosine similarity of word2vec embeddings to get similar words.
- Based on the assumption that, **semantically similar words often share the same hypernyms** (Qiu et. al. 2018).

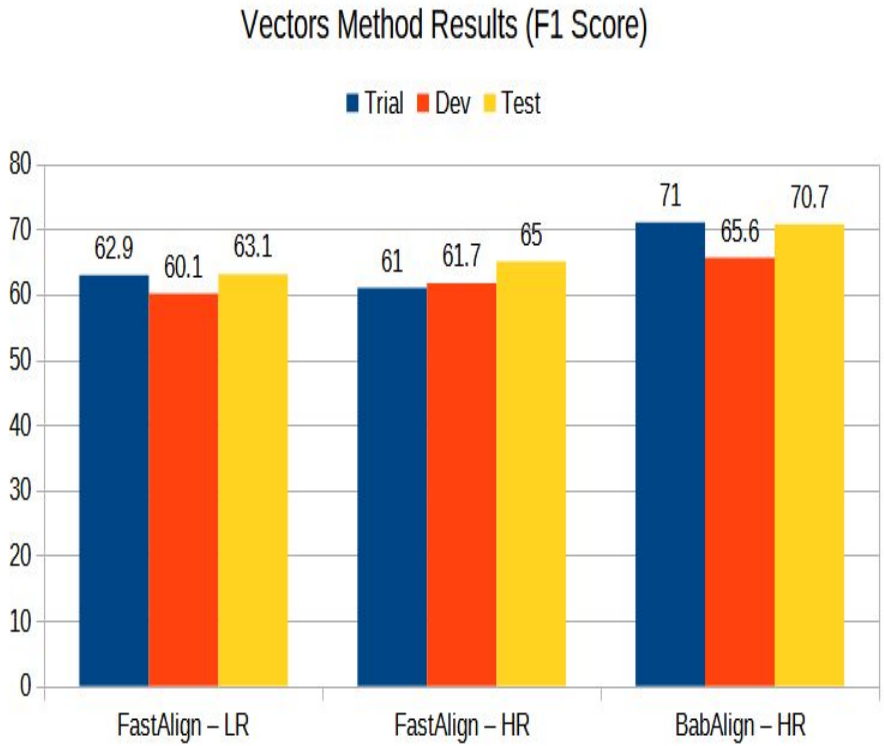


Experiments

- Bitext : OpenSubtitles → 22.5M aligned sentence pairs for EN-DE.
- Low & High Resource Setting :
  - Low Resource (LR) : 1M aligned sentences, no lemmatization
  - High Resource (HR) : full parallel corpora, lemmatization.
    - FastAlign
    - BabAlign
- Dataset : Train, Dev, Test sets from SemEval.

Results


Bitext Method Results (F1 Score)
Trial
Dev
Test
80
70
60
50
40
30
20
10
0
17.4
20.1
24.7
17
29
31.2
49.2
49.5
52.4
FastAlign – LR
FastAlign – HR
BabAlign – HR


Vectors Method Results (F1 Score)
Trial
Dev
Test
80
70
60
50
40
30
20
10
0
62.9
60.1
63.1
61
61.7
65
71
65.6
70.7
FastAlign – LR
FastAlign – HR
BabAlign – HR

Summary

- ➢ Results demonstrate a strong connection between translations and entailment.
- ➢ Weakness : methods unable to distinguish the direction of entailments. Leads to **false positive** cases. Example : creatura^{IT} does not entail wolf^{EN}, but our method can predict otherwise, if there is an entailment.

Recap

- Algorithms for correcting sense annotations.
  - Consistent improvements across all languages.
- Unsupervised corpus labelling approaches.
  - Achieved **state-of-the-art results** in multilingual unsupervised WSD.
- Translation based approaches for detecting entailment.
  - Demonstrated strong connection between entailment and translations.