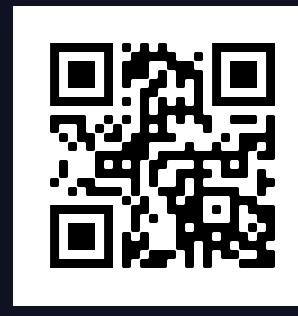




SENSEMBERT: CONTEXT-ENHANCED SENSE EMBEDDINGS FOR MULTILINGUAL WORD SENSE DISAMBIGUATION



BIANCA SCARLINI, TOMMASO PASINI AND ROBERTO NAVIGLI

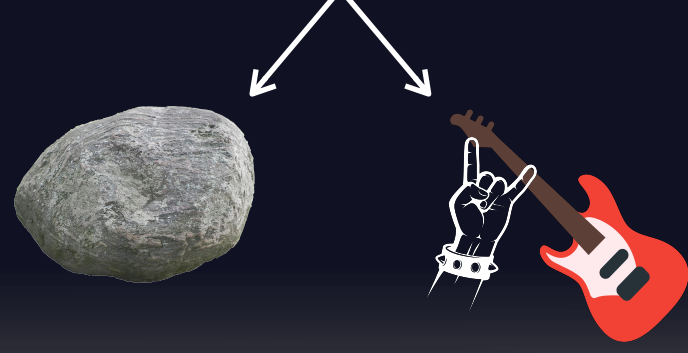


<http://sensembert.org>

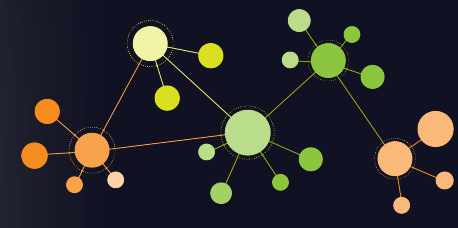


WORD SENSE DISAMBIGUATION

This guitar is good for playing **rock**.



CONTRIBUTIONS



Knowledge-based approach for building sense embeddings in **multiple languages**.

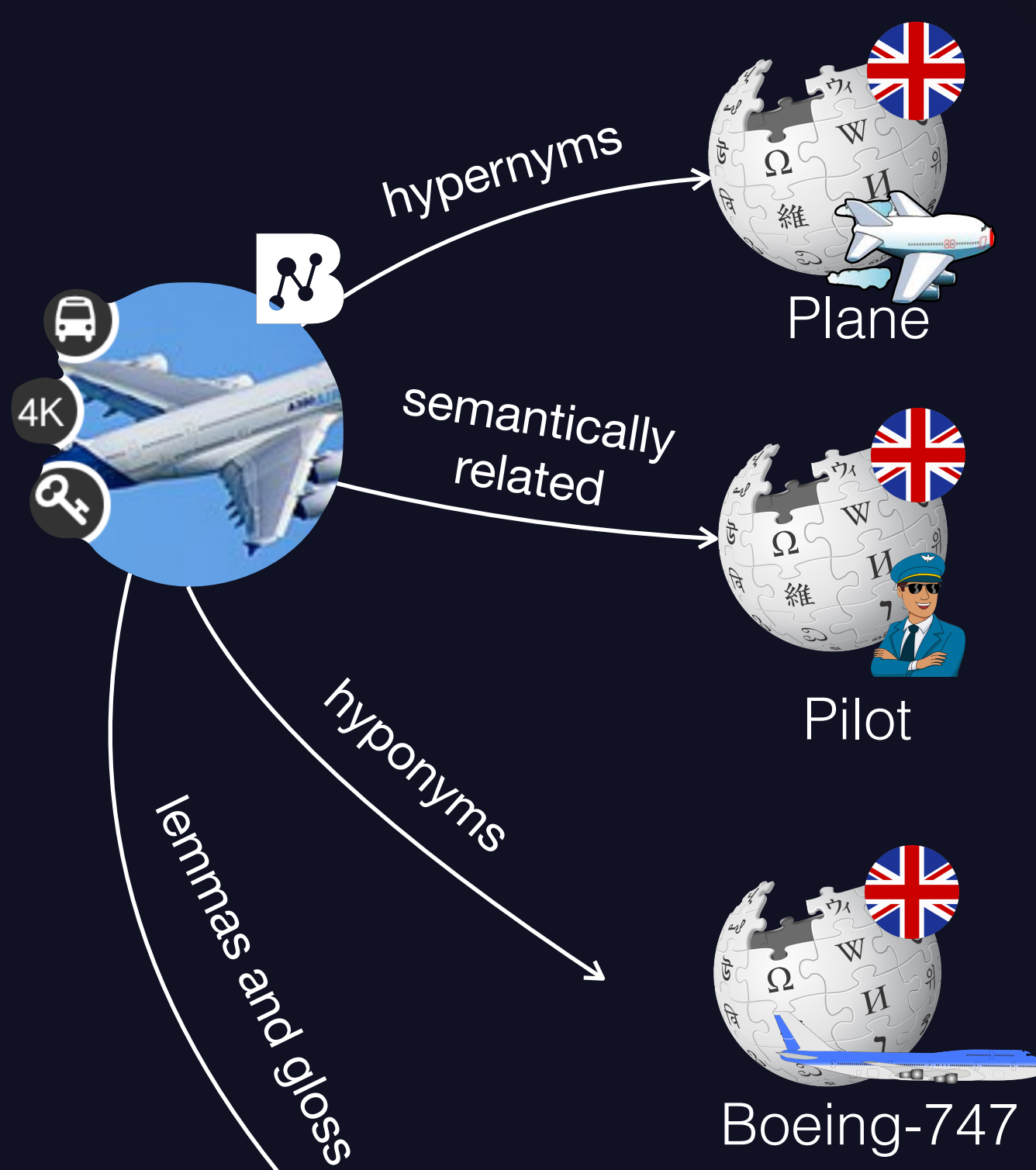
The embeddings take into account the **contexts** where a sense may appear.



The embeddings lay in a latent space that is **comparable** with the one of **BERT** contextual word embeddings.

METHOD

Relevant words



An **airplane** or **aeroplane** (informally **plane**) is a powered, **fixed-wing aircraft** that is **propelled** forward by thrust from a **jet engine**, propeller or **rocket engine**.

The number of **airline pilots** could decrease as **automation** replaces **copilots** and eventually **pilots** as well. [...] the number and kind of **flying jobs** available may be decreased.

The **quadjet 747** uses a **double-deck** configuration for part of its length and is available in **passenger**, freighter, and other versions. **Boeing** designed the **747**'s hump-like upper **deck** to serve as a **first-class lounge** or extra **seating**, and to allow the **aircraft** to be easily **converted** to a **cargo** carrier by removing **seats** and installing a front cargo door.

plane - plane, airplane. An aircraft that has a fixed wing and is powered by propellers or jets.

Context

~ airplane ~

~ aeroplane ~

~ plane ~

...

~ seats ~

~ plane ~

~ plane ~

~ airplane ~

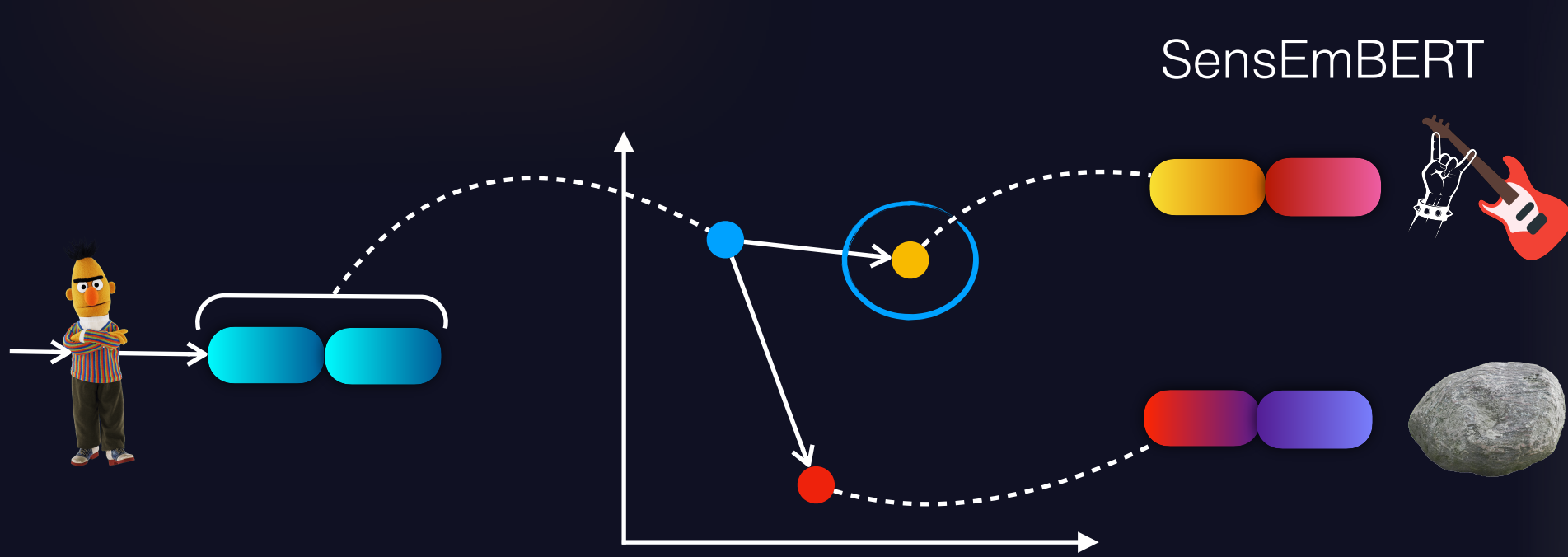
...

~ jets ~

Sense Gloss

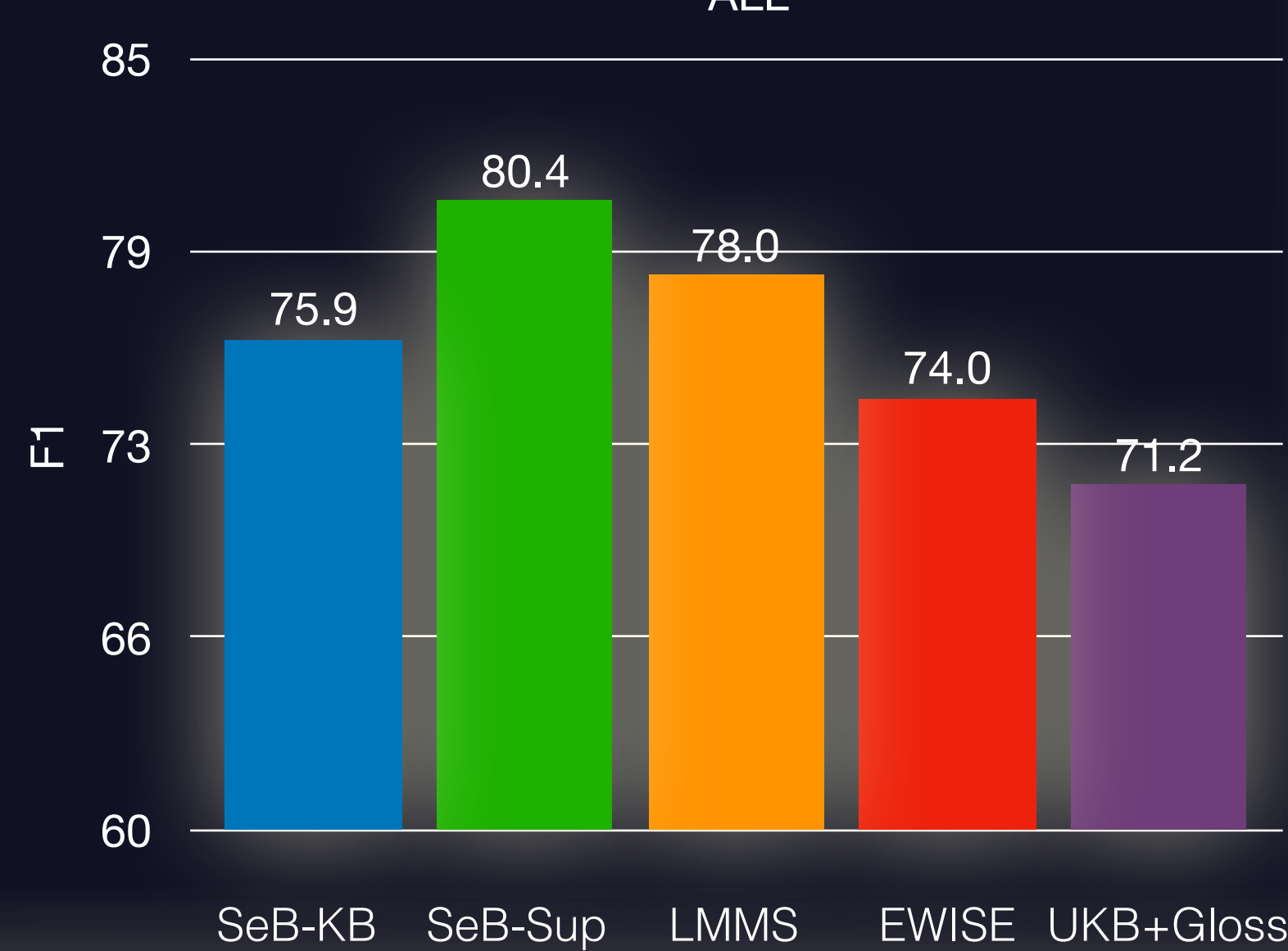
ENGLISH WSD

This guitar is good for playing **rock**.



Competing Systems:

SeB-KB: SensEmBERT embeddings.
SeB-Sup: Combines SemCor contexts with SeB-KB embeddings.
LMMS: Sense embeddings learned from SemCor.
EWISSE: Neural supervised approach.
UKB+Gloss: A knowledge-based algorithm exploiting glosses and semantic relations.



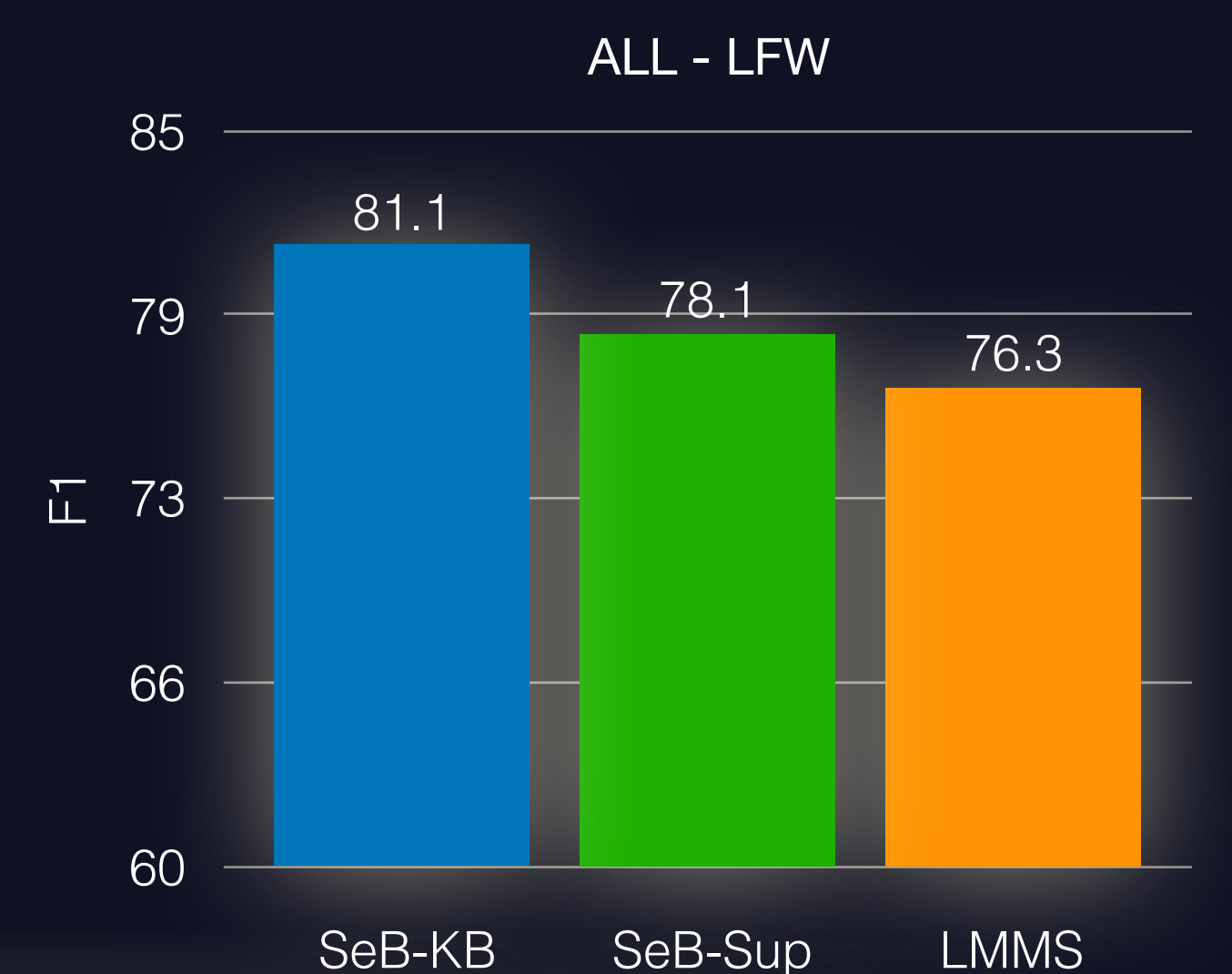
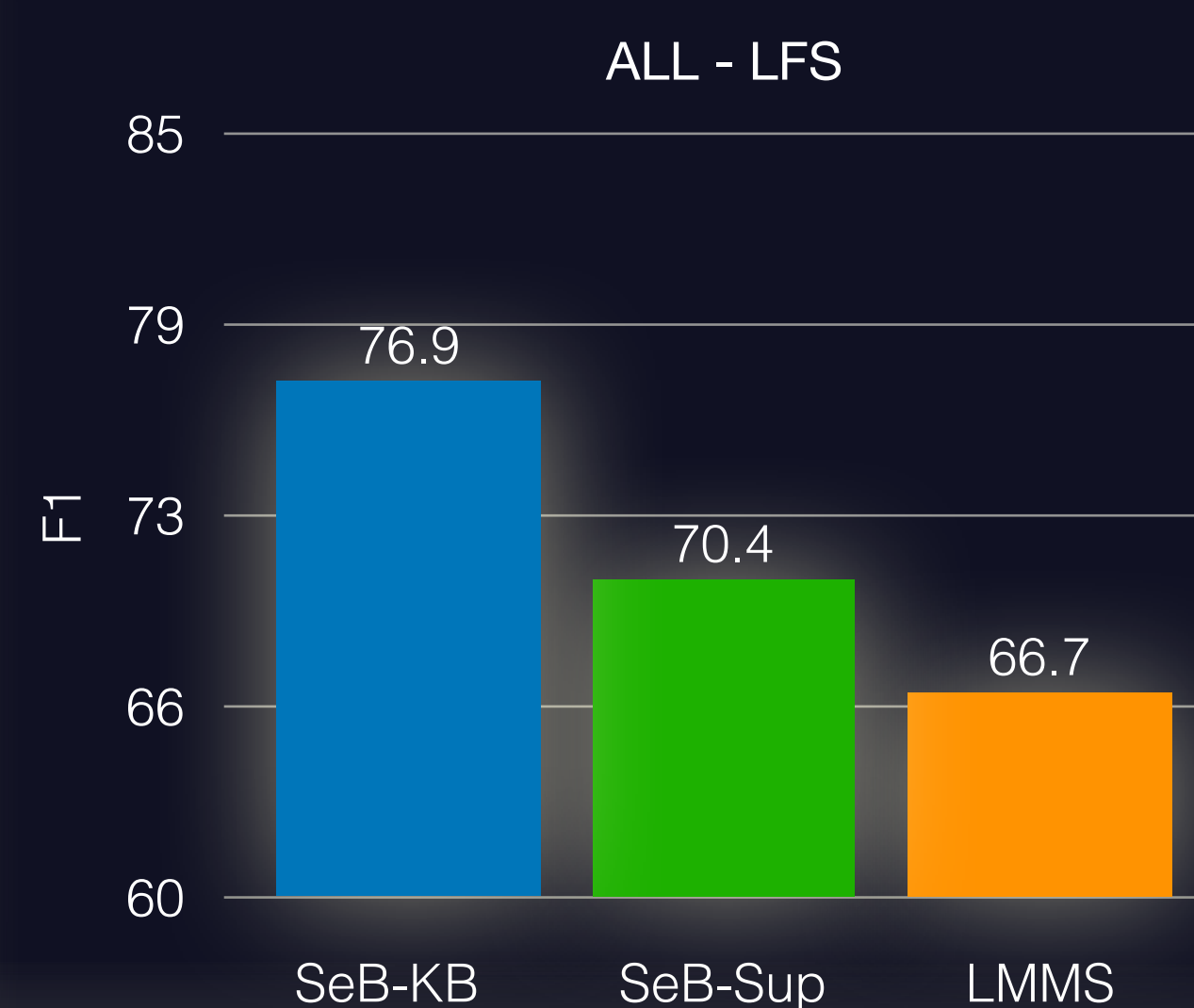
RARE INSTANCES

ALL - LFS (Least Frequent Senses):

Dataset containing all the instances in ALL tagged with a sense not in SemCor.

ALL - LFW (Least Frequent Words):

Dataset containing all the instances in ALL associated with a lemma not in SemCor.

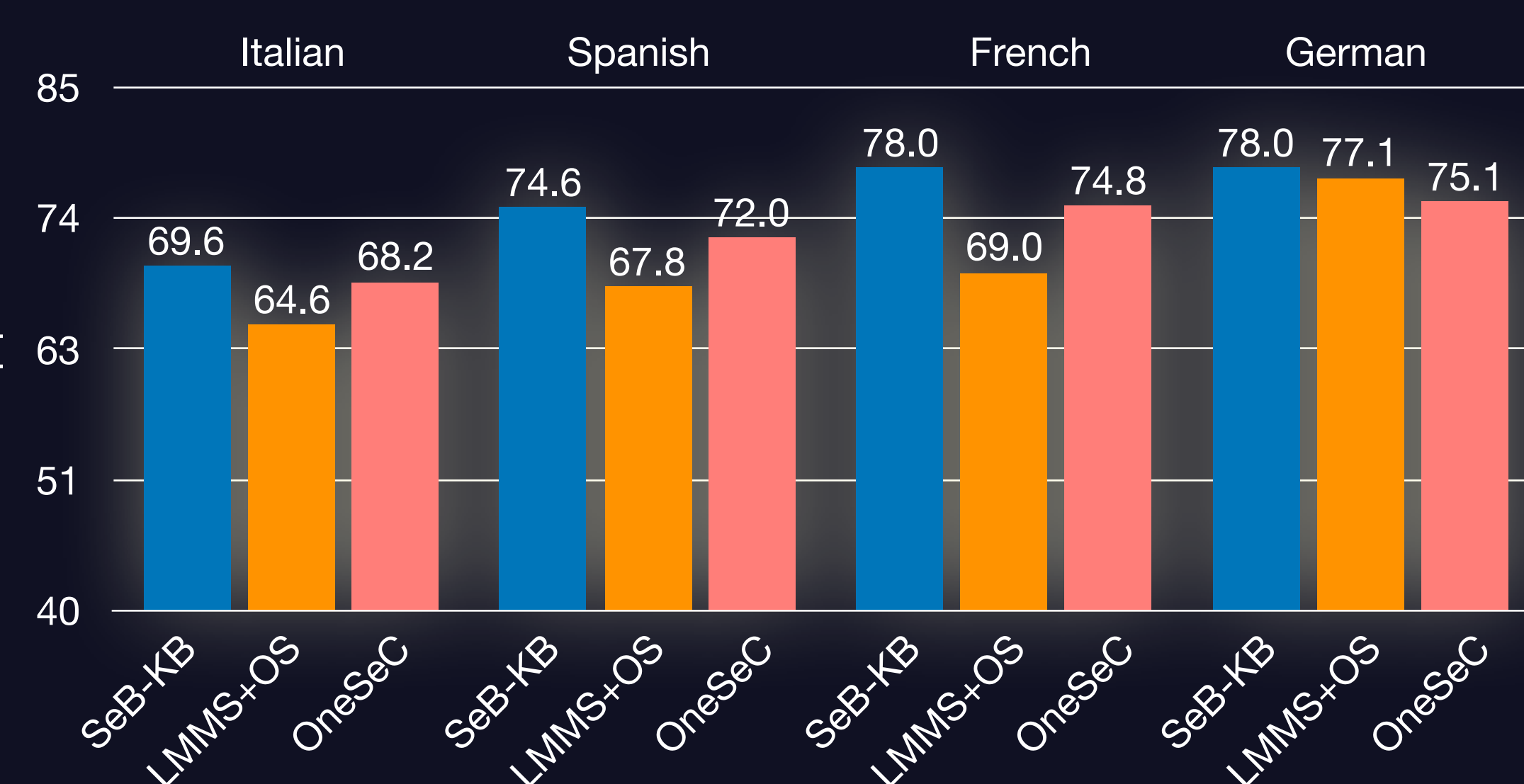


MULTILINGUAL WSD

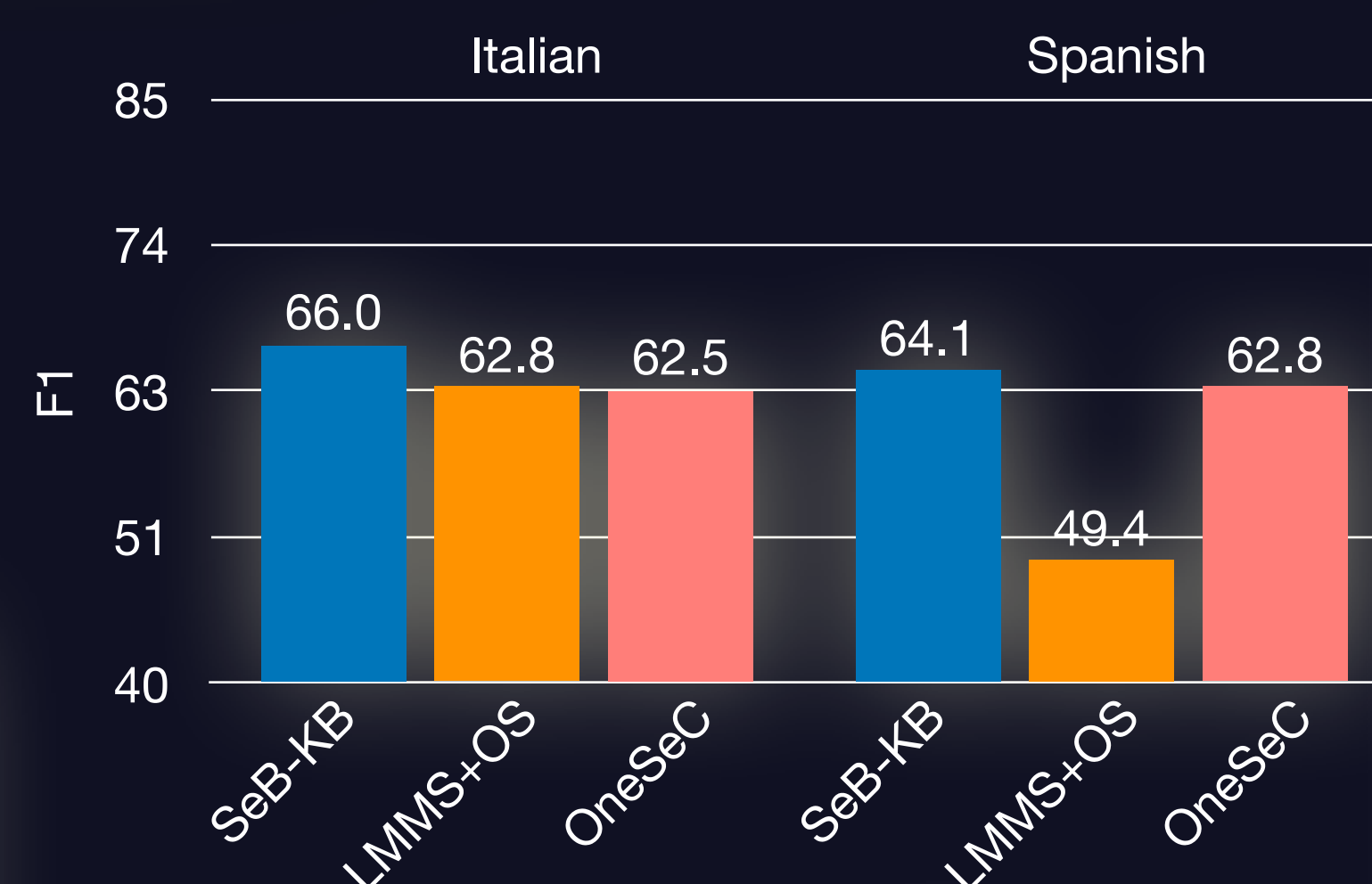
Competing Systems:

SeB-KB: SensEmBERT embeddings.
OneSeC: Neural supervised approach trained on OneSeC silver data.
LMMS+OS: Sense embeddings learned from OneSeC (OS).

SemEval-2013

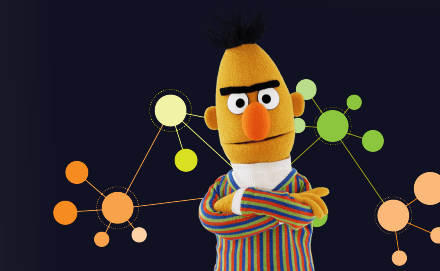


SemEval-2015



SOTA everywhere!

CONCLUSIONS & FUTURE WORKS



Knowledge-based approach for producing embeddings of senses in **multiple languages**.

SOTA results on the English and multilingual WSD tasks.

NOUNS
ADJECTIVES
VERBS
ADVERBS

Extend our approach to cover senses with **other POS tags**.



Supported by the ERC Consolidator Grant MOUSSE No. 726487 under the European Union's Horizon 2020 research and innovation programme

