SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual Word Sense Disambiguation

Bianca Scarlini, Tommaso Pasini and Roberto Navigli
Sapienza University of Rome
Department of Computer Science
{scarlini, pasini, navigli}@di.uniroma1.it

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

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual Word Sense Disambiguation

Bianca Scarlini, Tommaso Pasini and Roberto Navigli
Sapienza University of Rome
Department of Computer Science
{scarlini, pasini, navigli}@di.uniroma1.it

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

Word Sense Disambiguation

The task of associating a word in context with its meaning from a predefined inventory of senses

Word Sense Disambiguation

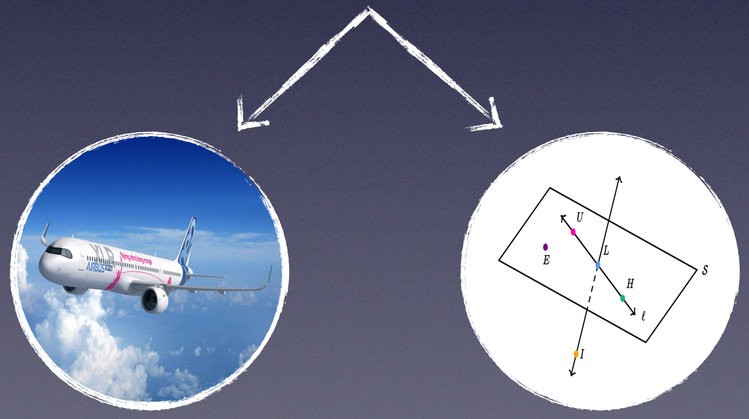
The task of associating a word in context with its meaning from a predefined inventory of senses

The **plane** just landed in New York.

Word Sense Disambiguation

The task of associating a word in context with its meaning from a predefined inventory of senses

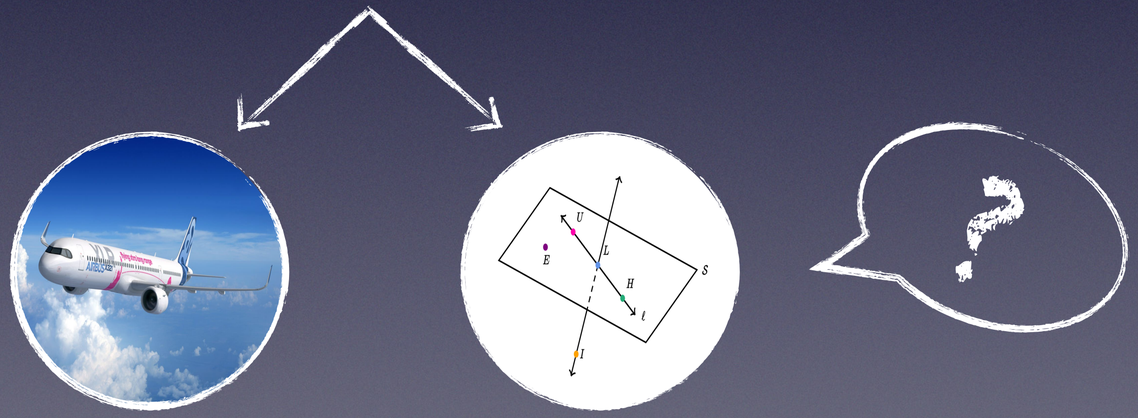
The **plane** just landed in New York.



Word Sense Disambiguation

The task of associating a word in context with its meaning from a predefined inventory of senses

The **plane** just landed in New York.



Supervised WSD

Knowledge-based WSD

Supervised WSD

NEURAL NETWORKS

SVM

ANNOTATED CORPORA

✓ Leverages machine-learning algorithms and sense-annotated training corpora.

Knowledge-based WSD

Supervised WSD

NEURAL NETWORKS

SVM

ANNOTATED CORPORA

✓ Leverages machine-learning algorithms and sense-annotated training corpora.

Knowledge-based WSD

KNOWLEDGE BASES

BAG OF WORDS

PERSONALISED PAGE RANK

✓ Leverages algorithms on knowledge and Bag of Words representation.

Supervised WSD

NEURAL NETWORKS

SVM

ANNOTATED CORPORA

✓ Leverages machine-learning algorithms and sense-annotated training corpora.

– *Hardly scales over rare words and languages.*

Knowledge-based WSD

KNOWLEDGE BASES

BAG OF WORDS

PERSONALISED PAGE RANK

✓ Leverages algorithms on knowledge and Bag of Words representation.

– *Falls behind supervised approaches on English.*

SensEmBERT

Knowledge-based approach for building sense embeddings in multiple languages

SensEmBERT

Knowledge-based approach for building sense embeddings in multiple languages

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

SensEmBERT

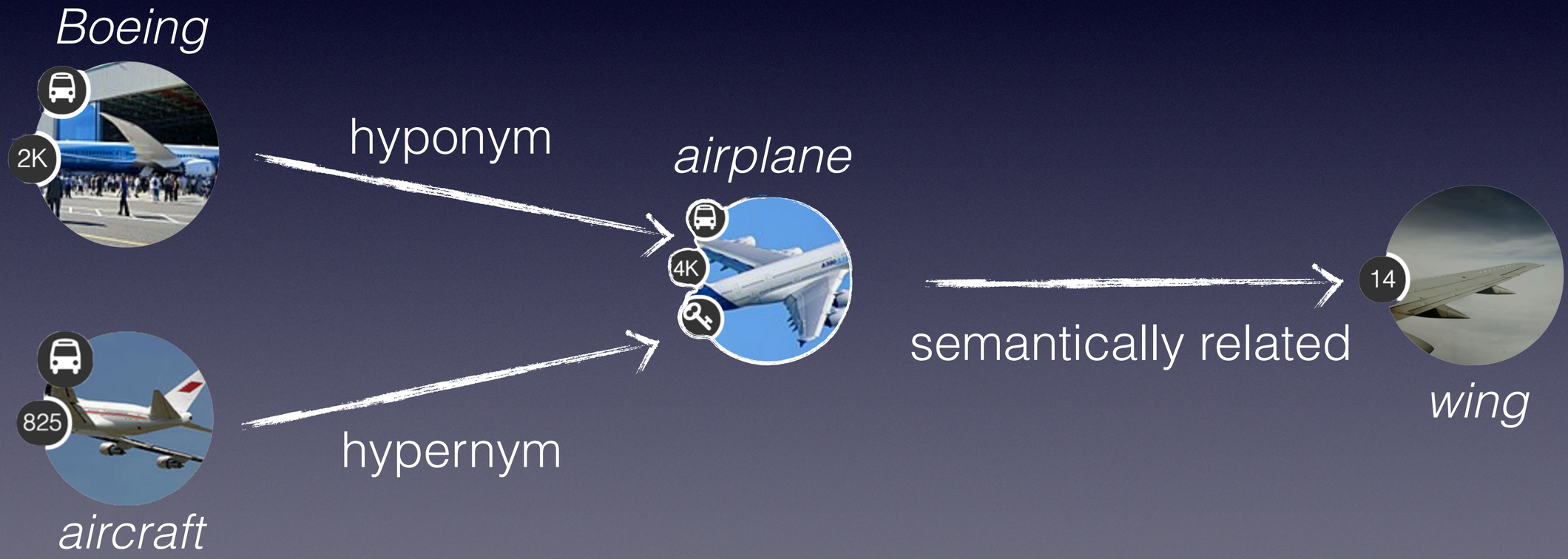
Knowledge-based approach for building sense embeddings in multiple languages

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

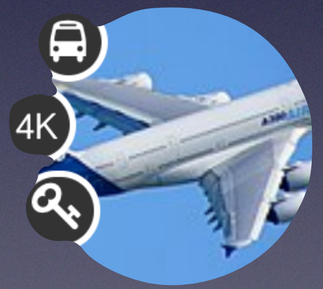
Preliminaries

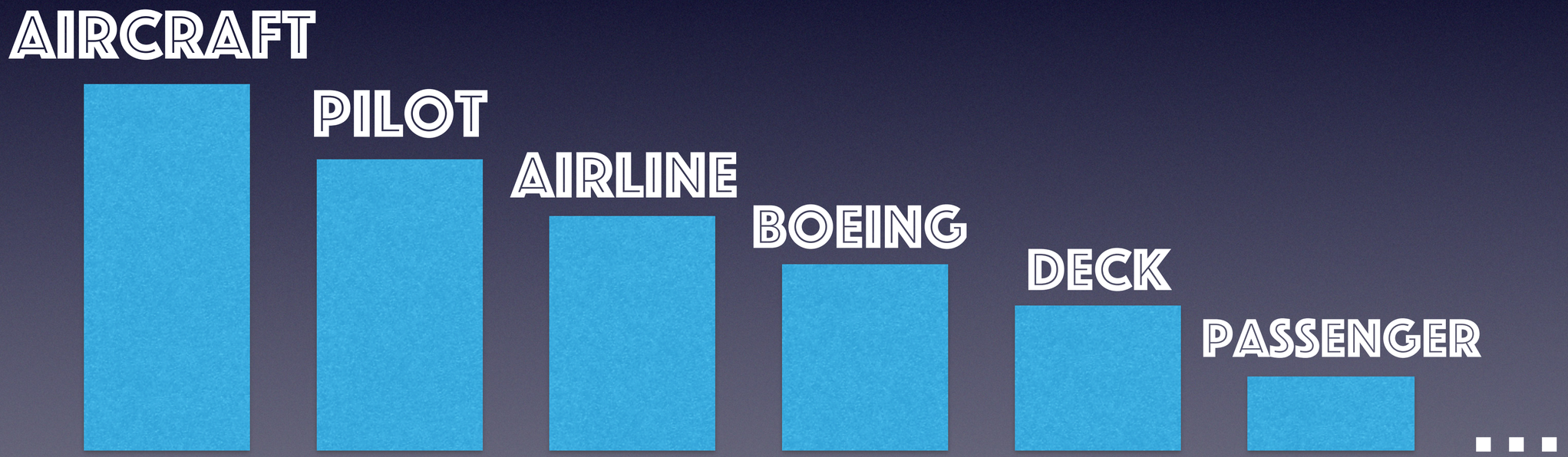
BabelNet is a multilingual semantic network



Preliminaries

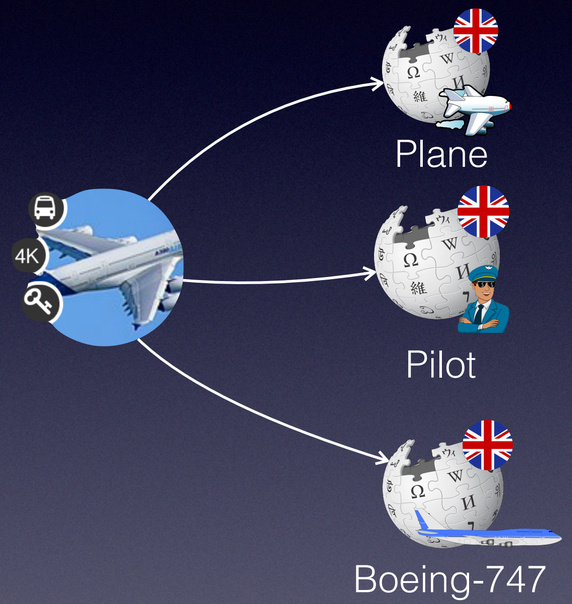
NASARI (Camacho-Collados et al. AIJ 2016) provides lexical **vector representations** of BabelNet **synsets**





SensEmBERT Overview

Related Pages



Plane

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.

Pilot

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.

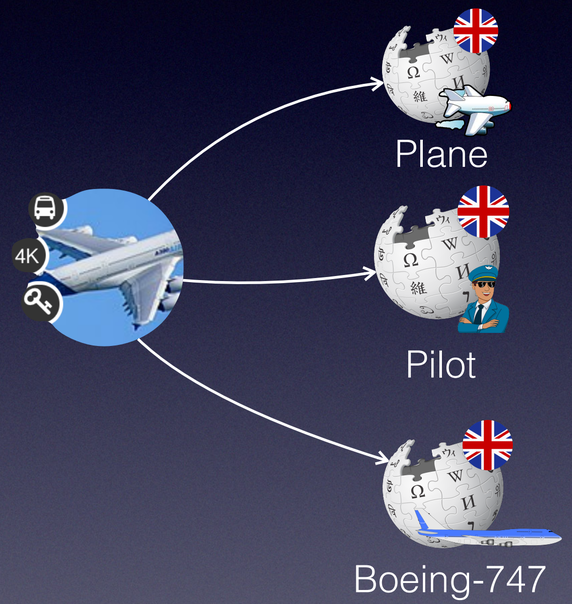
Boeing-747

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.

SensEmBERT Overview

STEP 2

Related Pages

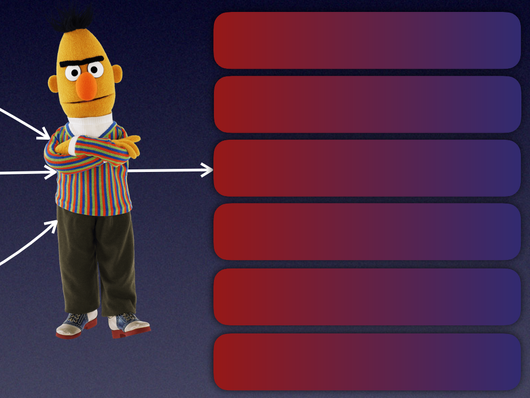


Relevant words

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**.

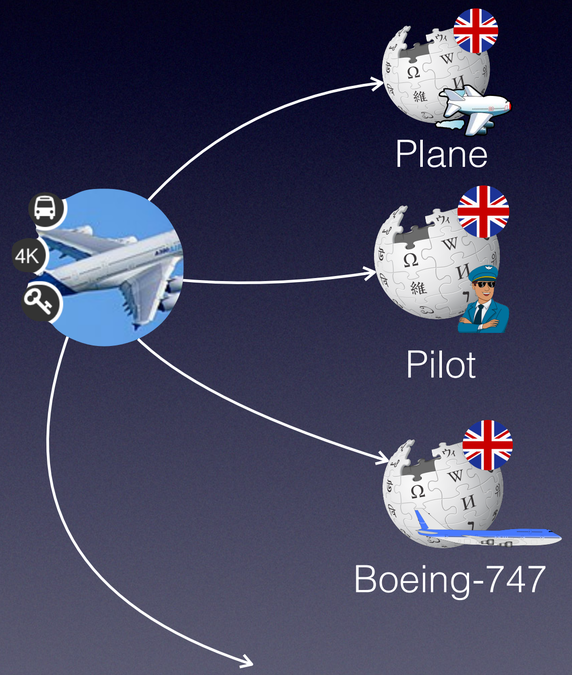


SensEmBERT Overview

STEP 3

Related Pages

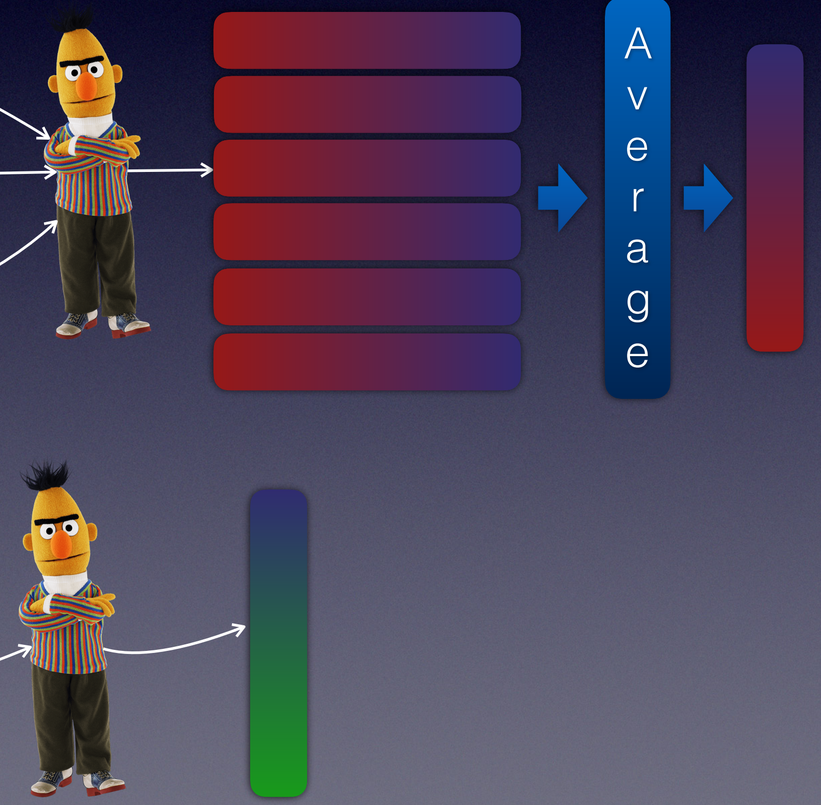
Relevant words



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.

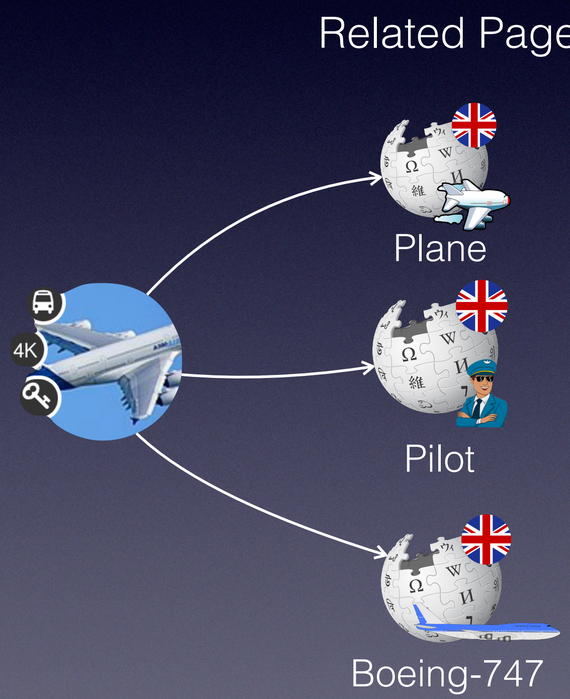
The **quadjet 747** uses a **double-deck** configuration for part of its length and is available in **passenger**, freighter, and other versions. **Boeing** designed the **747**'s hump-like upper **deck** to serve as a **first-class lounge** or extra **seating**, and to allow the **aircraft** to be easily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



Lemmas and Gloss

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

Step 1: Context Retrieval

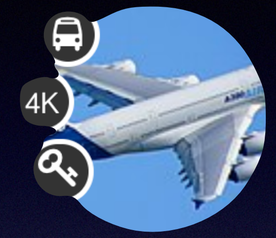


An airplane or aeroplane (informally plane) is a powered, fixed-wing aircraft that is propelled forward by thrust from a jet engine, propeller or rocket engine.

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.

Step 1: Context Retrieval



Hypernyms

Aircraft

Powered Aircraft

Fixed-wing Aircraft

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval

Hypernyms

Aircraft

Powered Aircraft

Fixed-wing Aircraft

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval

Hypernyms

Aircraft

Powered Aircraft

Fixed-wing Aircraft

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

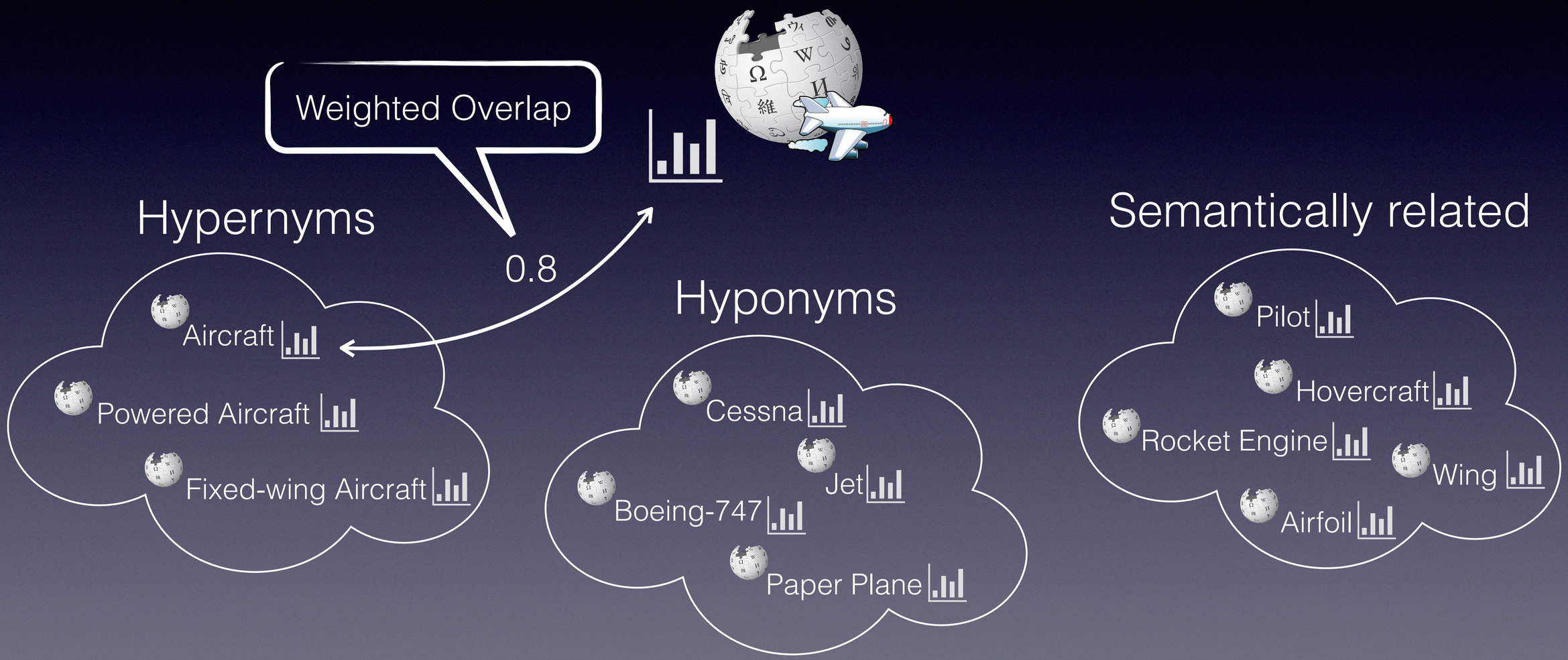
Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval



Step 1: Context Retrieval

Hypernyms

Aircraft ***0.8***

Powered Aircraft

Fixed-wing Aircraft

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval

Hypernyms

Aircraft ***0.8***

Powered Aircraft ***0.6***

Fixed-wing Aircraft

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval

Hypernyms

Aircraft ***0.8***

Powered Aircraft ***0.6***

Fixed-wing Aircraft ***0.7***

Hyponyms

Cessna

Jet

Boeing-747

Paper Plane

Semantically related

Pilot

Hovercraft

Rocket Engine

Wing

Airfoil

Step 1: Context Retrieval

Hypernyms

Aircraft ***0.8***

Powered Aircraft ***0.6***

Fixed-wing Aircraft ***0.7***

Hyponyms

Cessna ***0.7***

Jet ***0.9***

Boeing-747 ***0.9***

Paper Plane ***0.5***

Semantically related

Pilot ***0.8***

Hovercraft ***0.4***

Rocket Engine ***0.6***

Wing ***0.9***

Airfoil ***0.7***

Step 1: Context Retrieval

Hypernyms

Aircraft ***0.8***

Powered Aircraft ***0.6***

Fixed-wing Aircraft ***0.7***

Hyponyms

Cessna ***0.7***

Jet ***0.9***

Boeing-747 ***0.9***

Paper Plane ***0.5***

Semantically related

Pilot ***0.8***

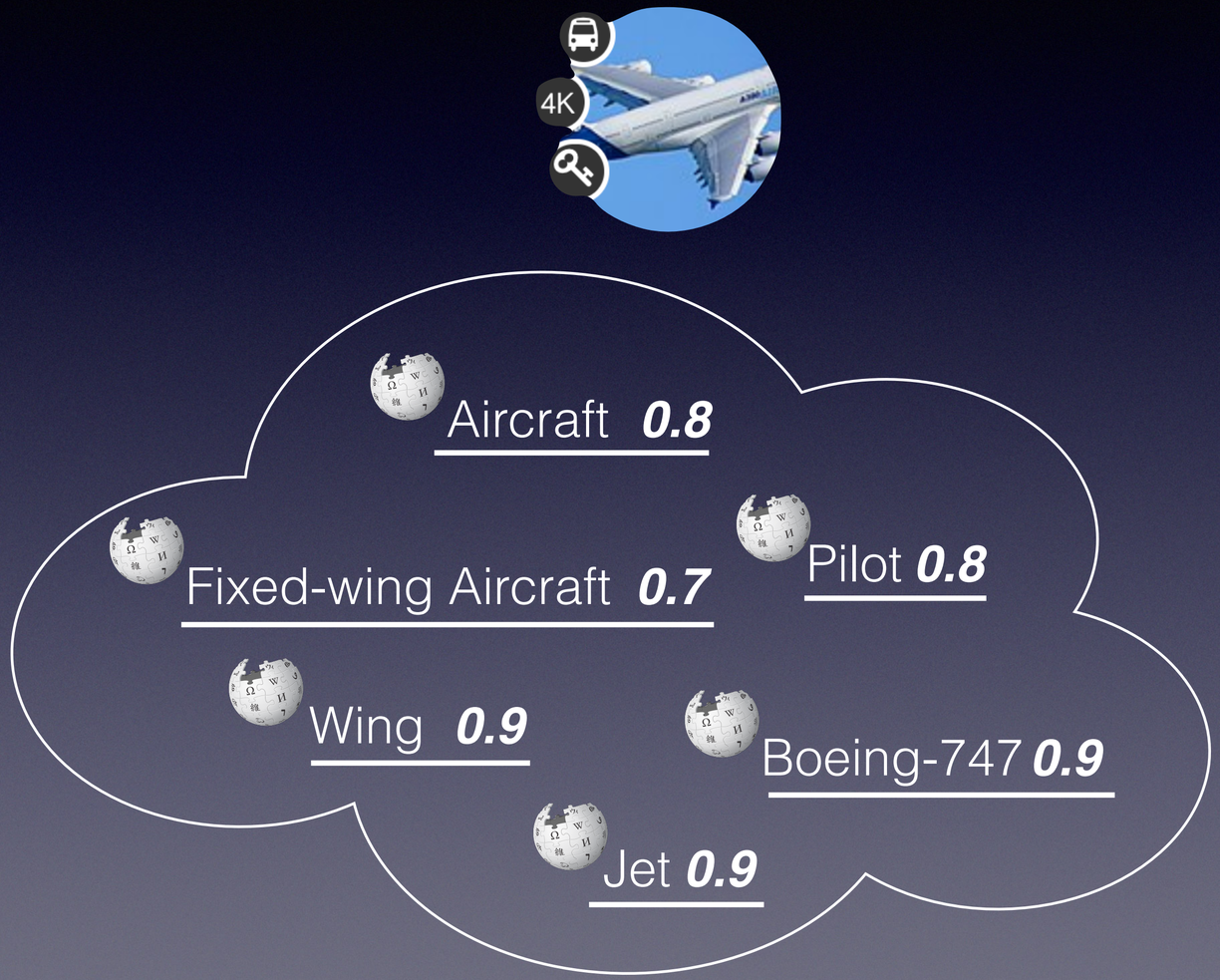
Hovercraft ***0.4***

Rocket Engine ***0.6***

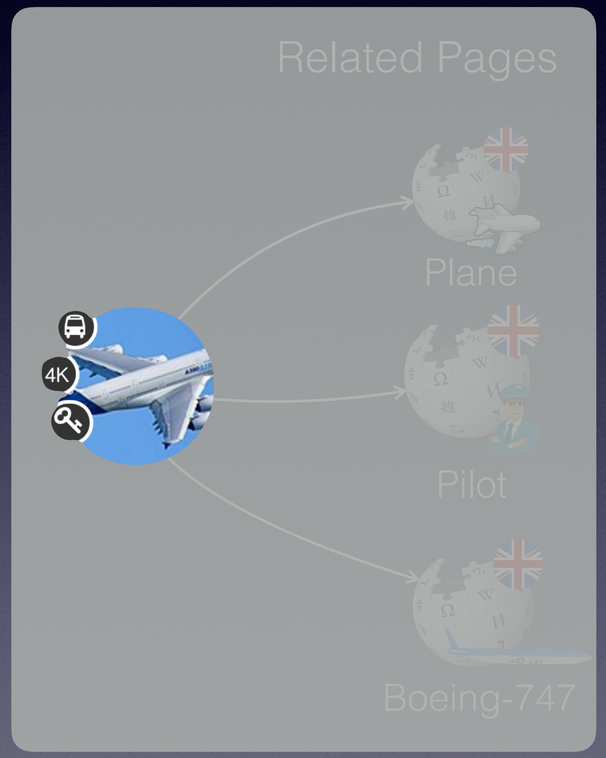
Wing ***0.9***

Airfoil ***0.7***

Step 1: Context Retrieval



Step 2: Word Embedding

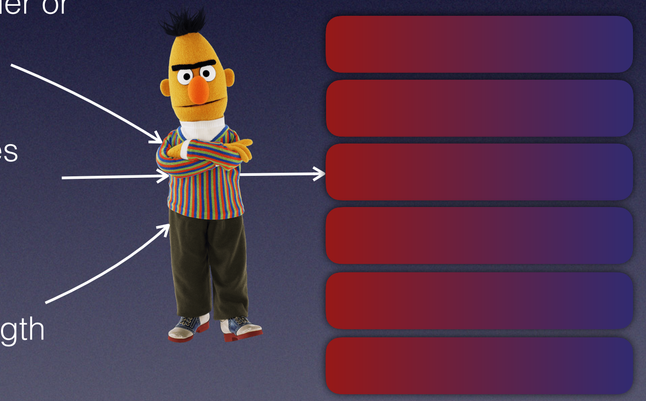


Relevant words

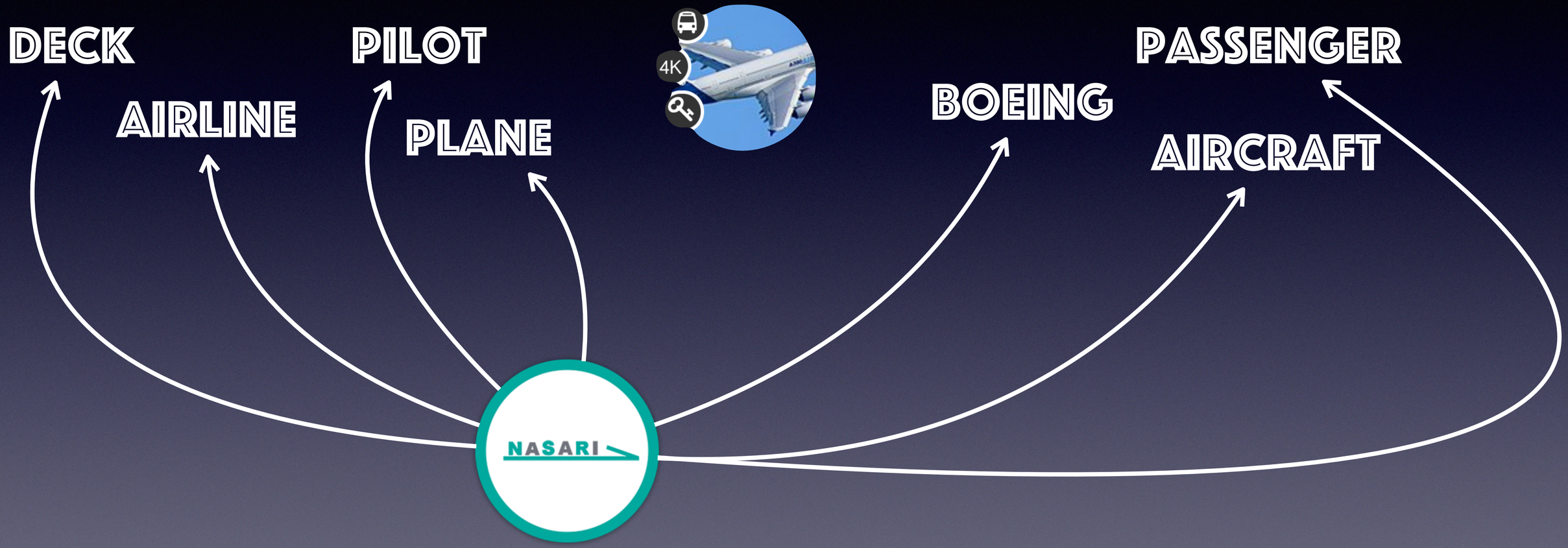
An **airplane** or **aeroplane** (informally **plane**) is a powered, **fixed-wing aircraft** that is **propelled** forward by thrust from a **jet engine**, propeller or **rocket engine**.

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**.



Step 2: Word Embedding



Step 2: Word Embedding

DECK

PILOT

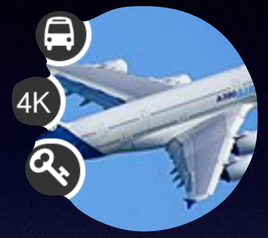
PASSENGER

AIRLINE

PLANE

BOEING

AIRCRAFT



Aircraft ***0.8***

Pilot ***0.8***

Fixed-wing Aircraft ***0.7***

Boeing-747 ***0.9***

Wing ***0.9***

Jet ***0.9***

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

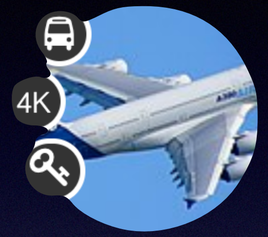
Step 2: Word Embedding

DECK

AIRLINE

PILOT

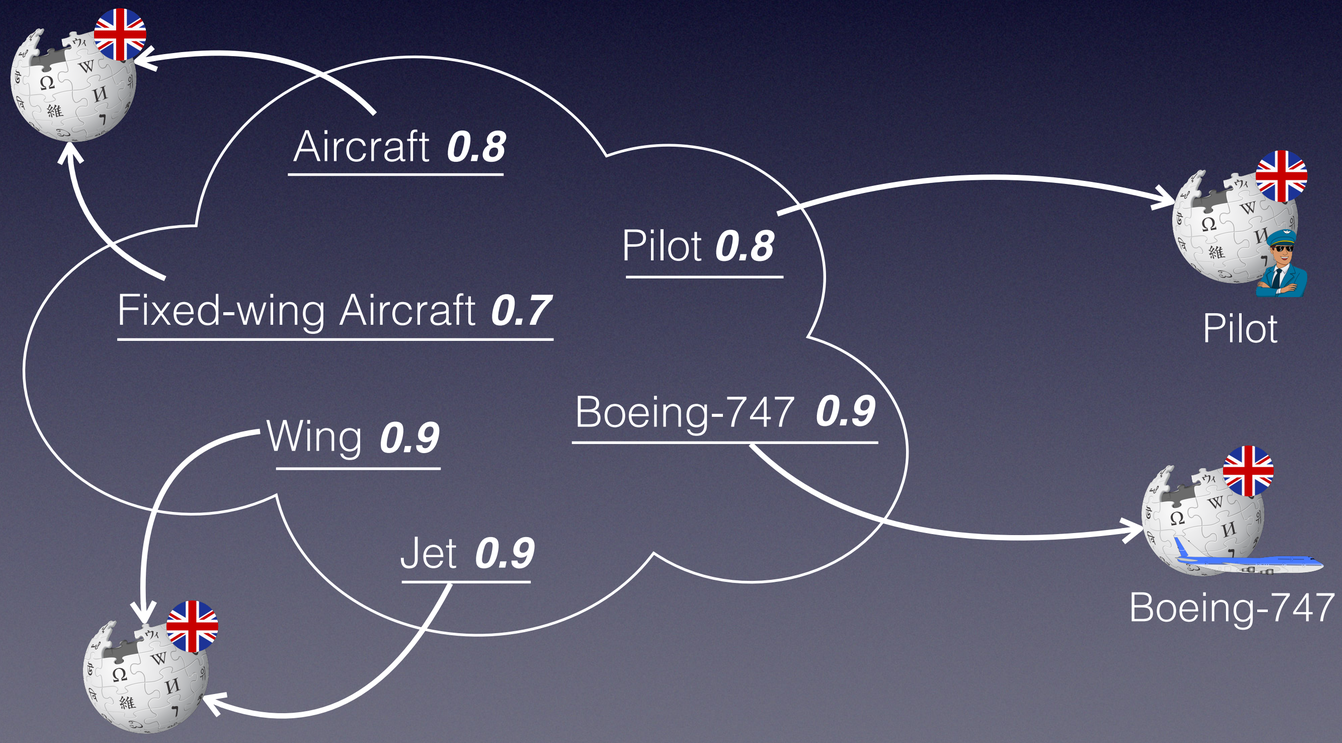
PLANE



PASSENGER

BOEING

AIRCRAFT



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.

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

Step 2: Word Embedding

DECK

PILOT

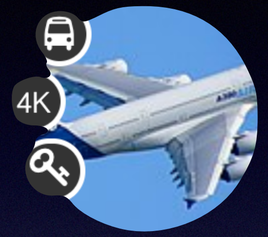
PASSENGER

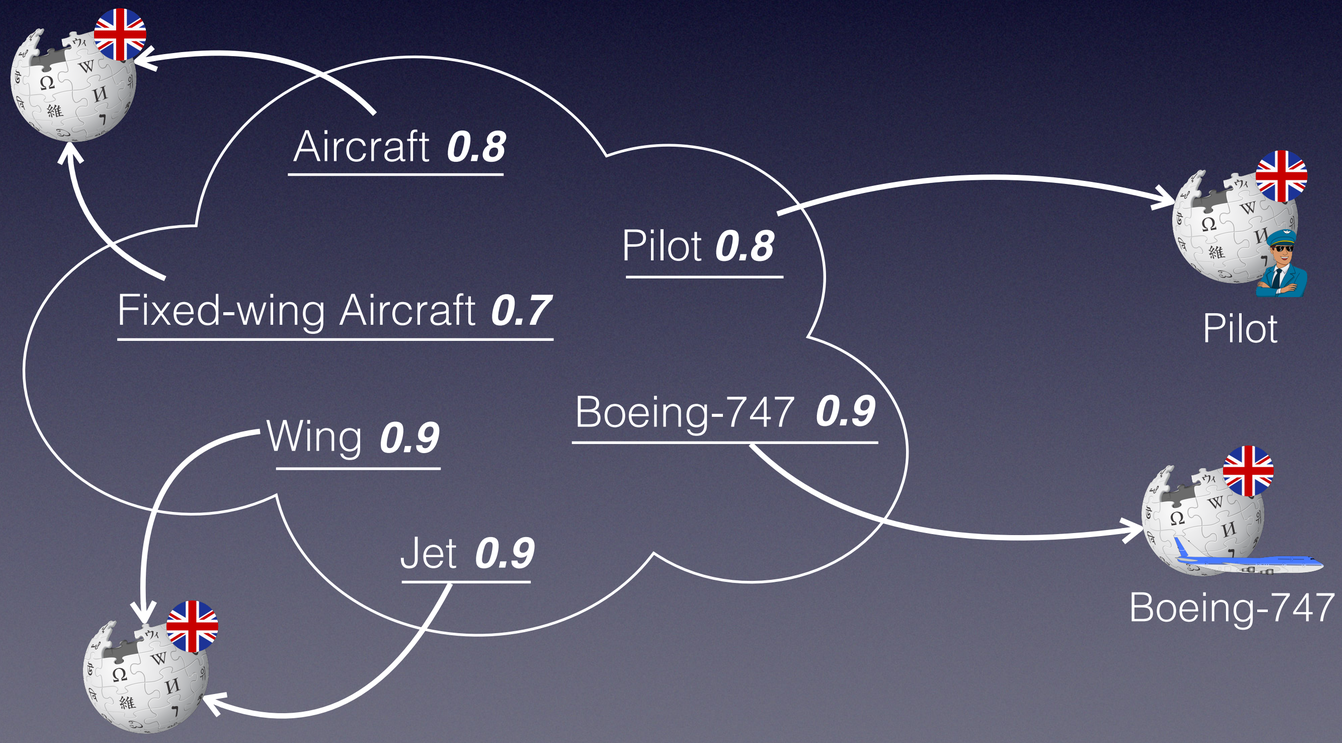
AIRLINE

PLANE

BOEING

AIRCRAFT





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.

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

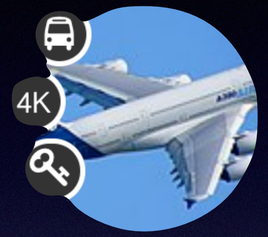
Step 2: Word Embedding

DECK

PILOT

AIRLINE

PLANE



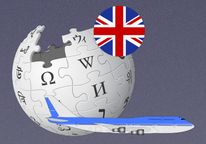
BOEING

PASSENGER

AIRCRAFT

Pilot

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.



Boeing-747

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.

airline

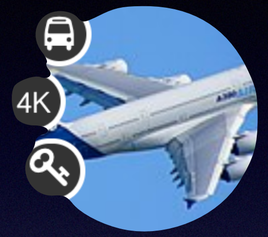
Step 2: Word Embedding

DECK

PILOT

AIRLINE

PLANE



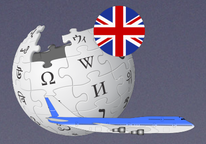
BOEING

PASSENGER

AIRCRAFT

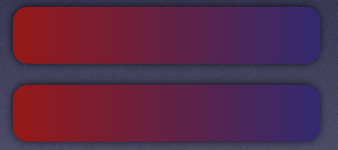
Pilot

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.



Boeing-747

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.



airline

pilots

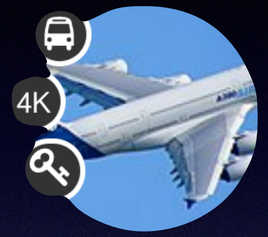
Step 2: Word Embedding

DECK

AIRLINE

PILOT

PLANE



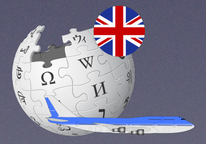
BOEING

PASSENGER

AIRCRAFT

Pilot

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.



Boeing-747

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.

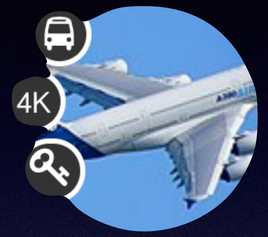
Step 2: Word Embedding

DECK

AIRLINE

PILOT

PLANE



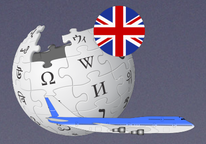
BOEING

PASSENGER

AIRCRAFT

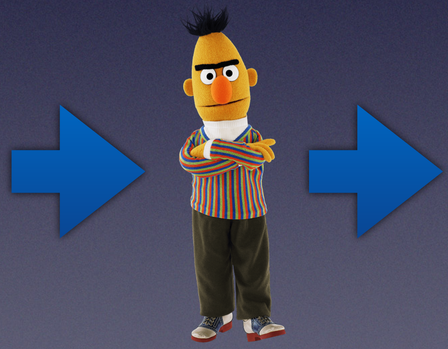
Pilot

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.



Boeing-747

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.



airline

pilots

pilots

deck

passenger

Boeing

deck

aircraft

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

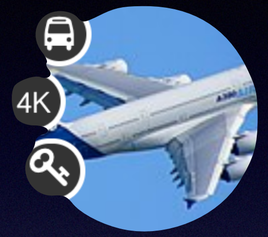
Step 2: Word Embedding

DECK

AIRLINE

PILOT

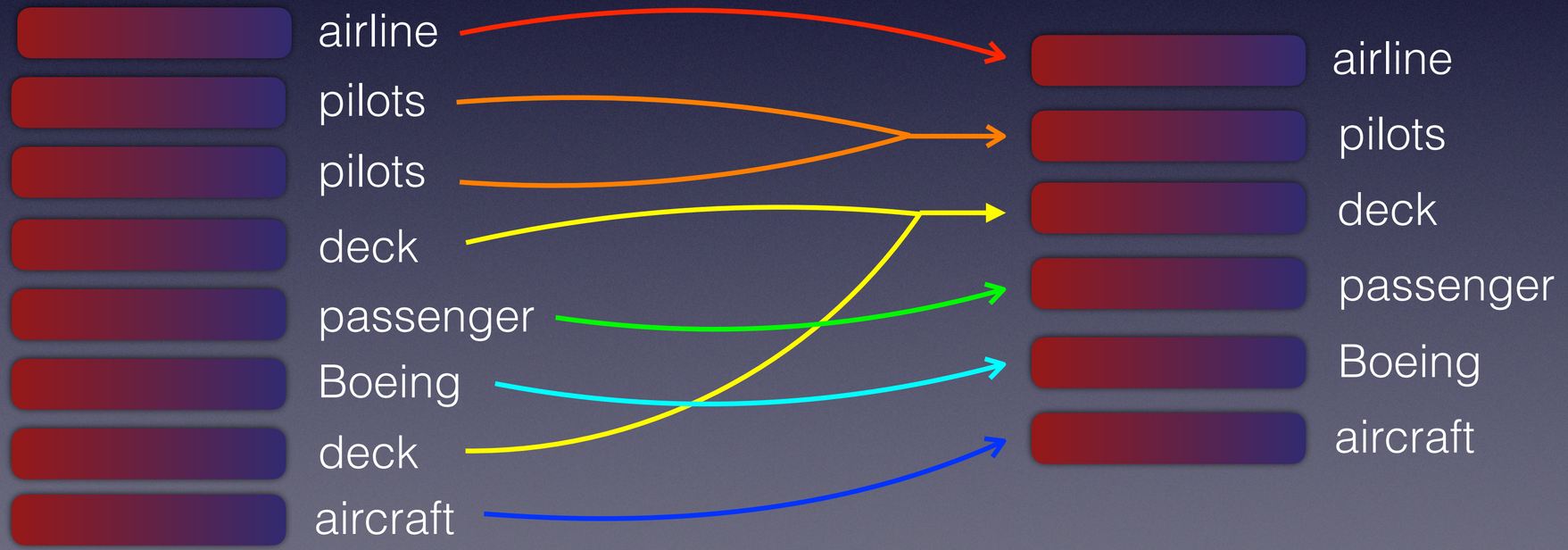
PLANE



BOEING

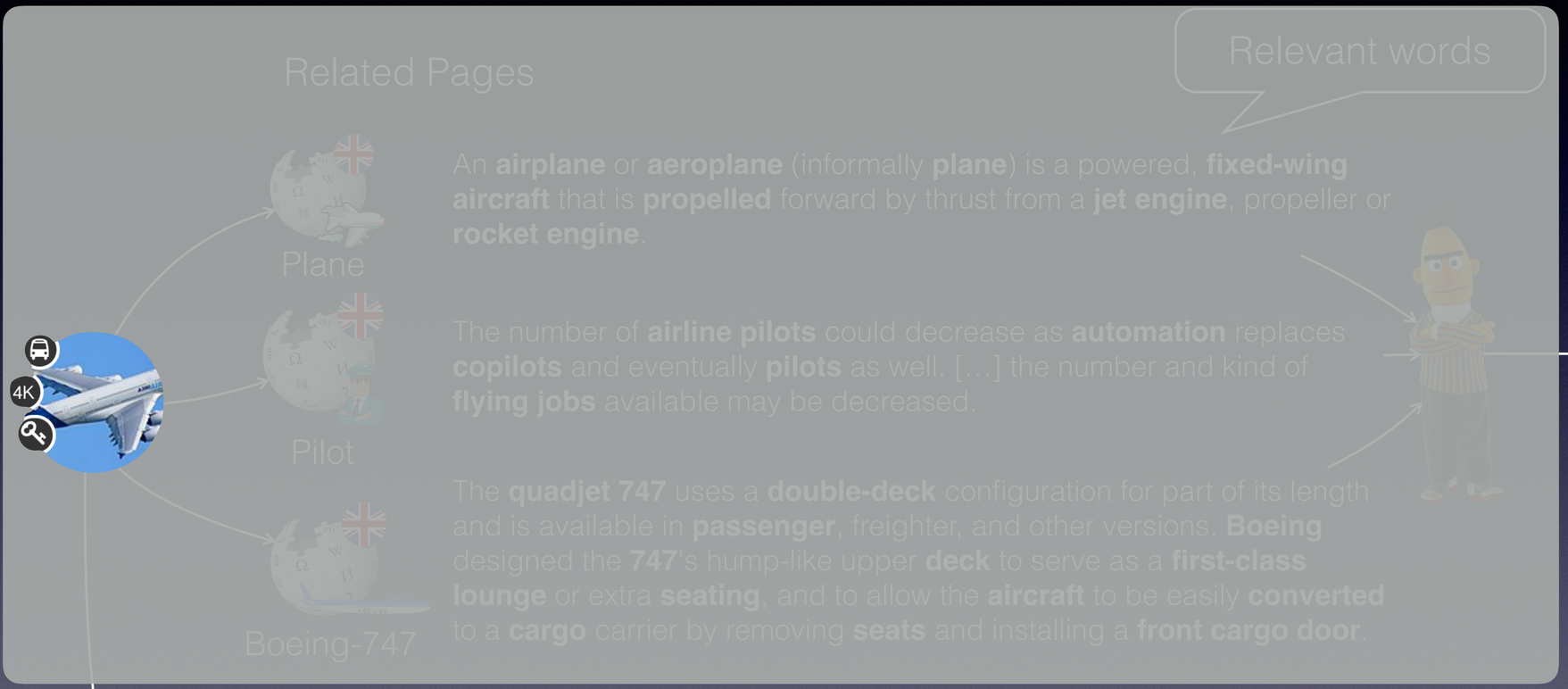
PASSENGER

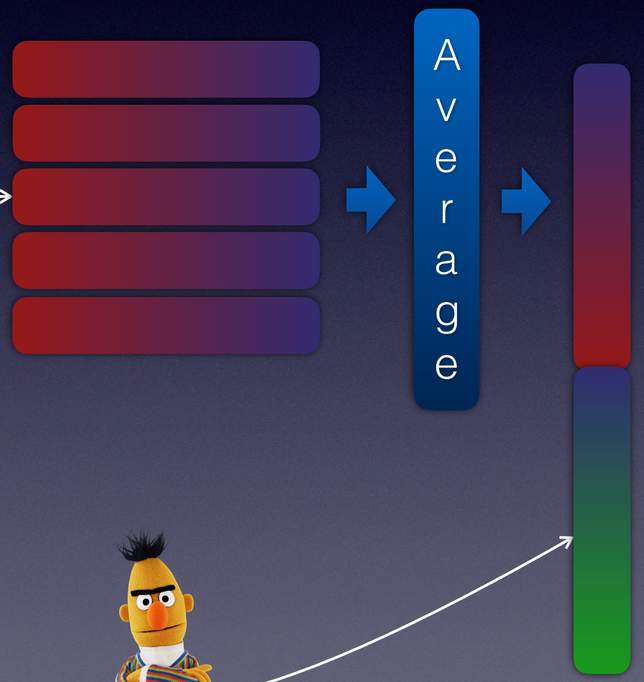
AIRCRAFT



SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

Step 3: Sense Embedding

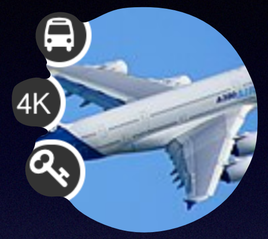


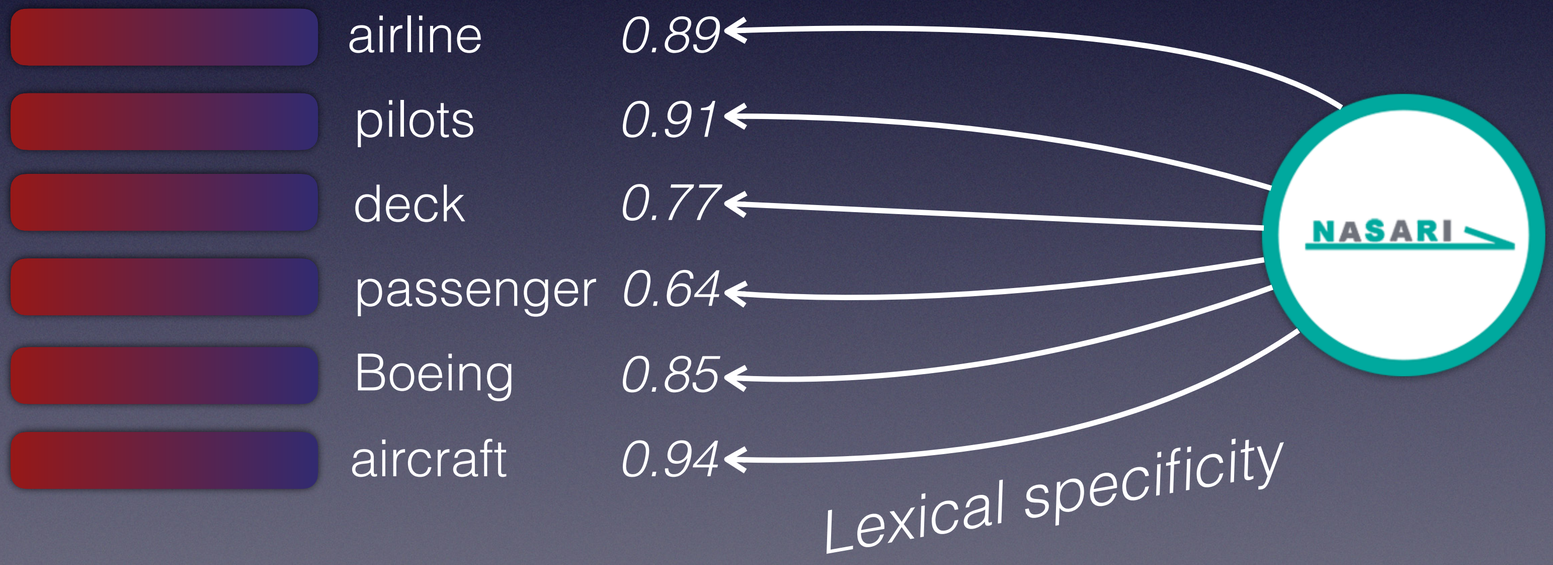


Lemmas and Gloss

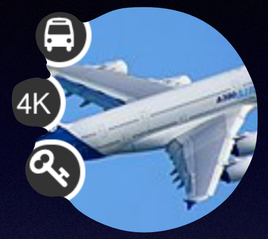
plane - plane, airplane. An aircraft that has a fixed wing and is powered by propellers or jets.

Step 3: Sense Embedding



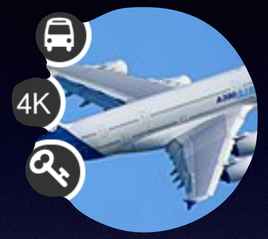


Step 3: Sense Embedding



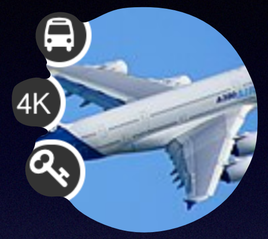
airline	*0.89*
pilots	*0.91*
deck	*0.77*
passenger	*0.64*
Boeing	*0.85*
aircraft	*0.94*

Step 3: Sense Embedding



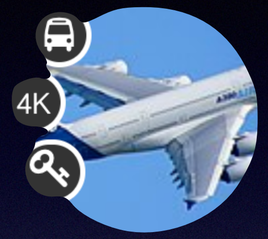
aircraft	*0.94*
pilots	*0.91*
airline	*0.89*
Boeing	*0.85*
deck	*0.77*
passenger	*0.64*

Step 3: Sense Embedding



	aircraft	*1*
	pilots	*2*
	airline	*3*
	Boeing	*4*
	deck	*5*
	passenger	*6*

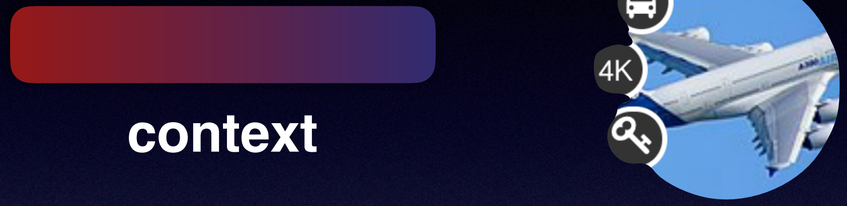
Step 3: Sense Embedding



	aircraft	*1*
	pilots	*2*
	airline	*3*
	Boeing	*4*
	deck	*5*
	passenger	*6*

context

Step 3: Sense Embedding



context

plane, airplane

lemmas

BabelNet

An aircraft that has a fixed wing and is powered by propellers or jets.

Step 3: Sense Embedding

context

lemmas

An aircraft that has a fixed wing and is powered by propellers or jets.

Step 3: Sense Embedding

context **lemmas**

An aircraft that has a fixed wing and is powered by propellers or jets.

prepend all the lemmas of the target synset

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

Step 3: Sense Embedding

context

lemmas

An aircraft that has a fixed wing and is powered by propellers or jets.

prepend all the lemmas of the target synset

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

Step 3: Sense Embedding

An aircraft that has a fixed wing and is powered by propellers or jets.

prepend all the lemmas of the target synset

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

repeat the target lemma

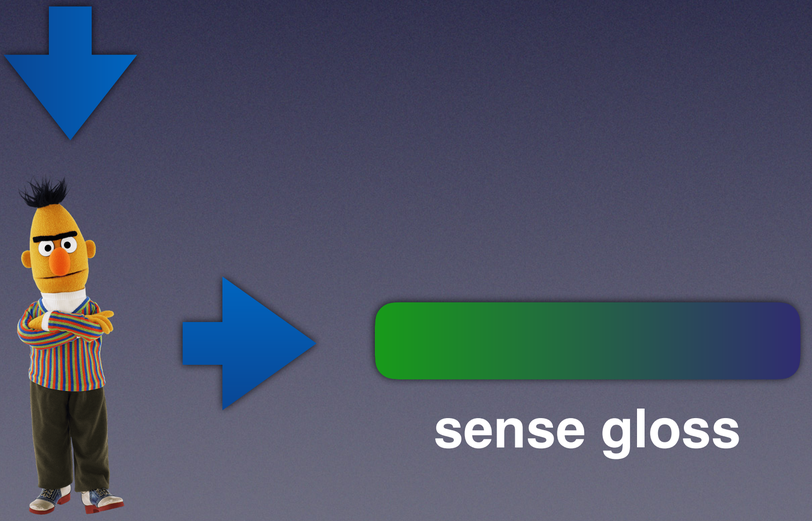
plane - plane, airplane - An aircraft that has a fixed wing and is powered by propellers or jets.

Step 3: Sense Embedding

context

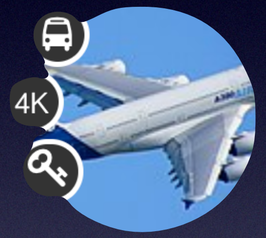
lemmas

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



sense gloss

Step 3: Sense Embedding



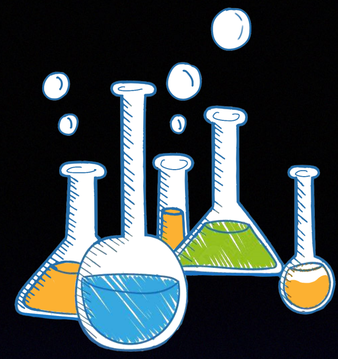
plane

airplane

context sense gloss

context sense gloss

Experimental Setup

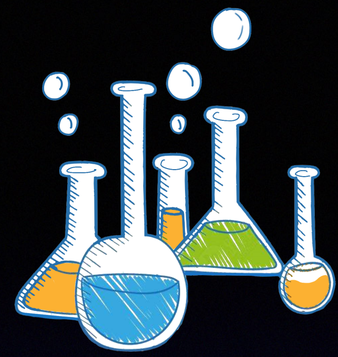


SensEmBERT WSD Model

- How does SensEmBERT disambiguate?

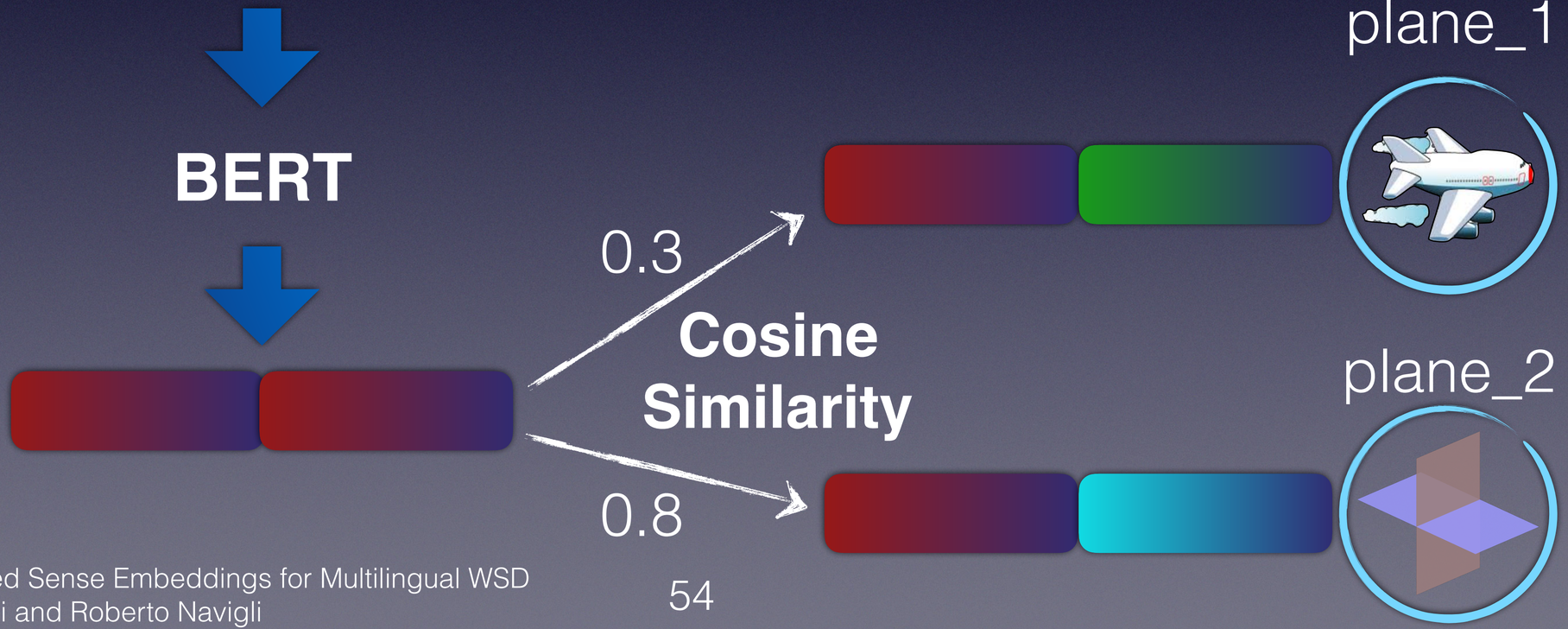
  Two distinct **planes** are either parallel or they intersect in a line.

Experimental Setup

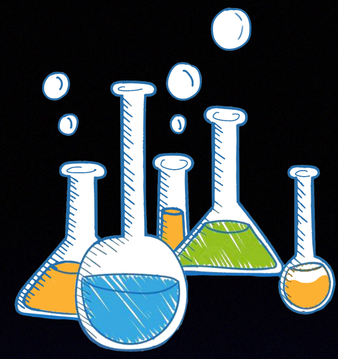


SensEmBERT WSD Model

Two distinct **planes** are either parallel or they intersect in a line.

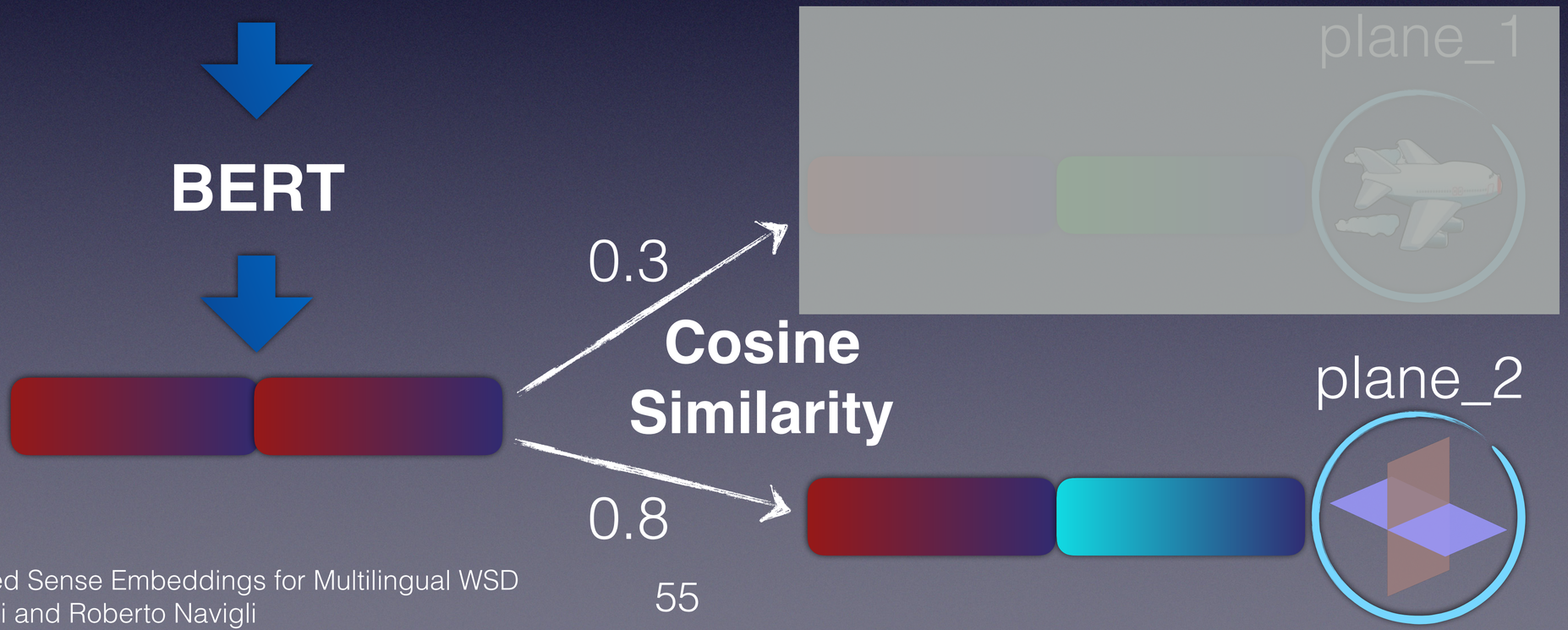


Experimental Setup



SensEmBERT WSD Model

Two distinct **planes** are either parallel or they intersect in a line.



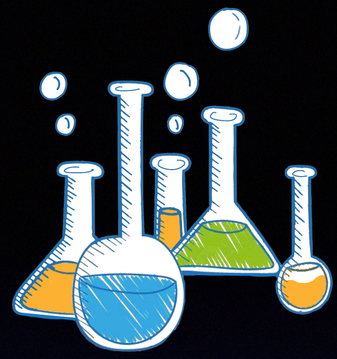
SensEmBERT Models

1) Knowledge-Based (SeB-KB): We consider as context to build a sense vector all those gathered as described before.

2) Supervised (SeB-Sup): Additionally, we consider as contexts for a sense all those in SemCor where the sense appears.

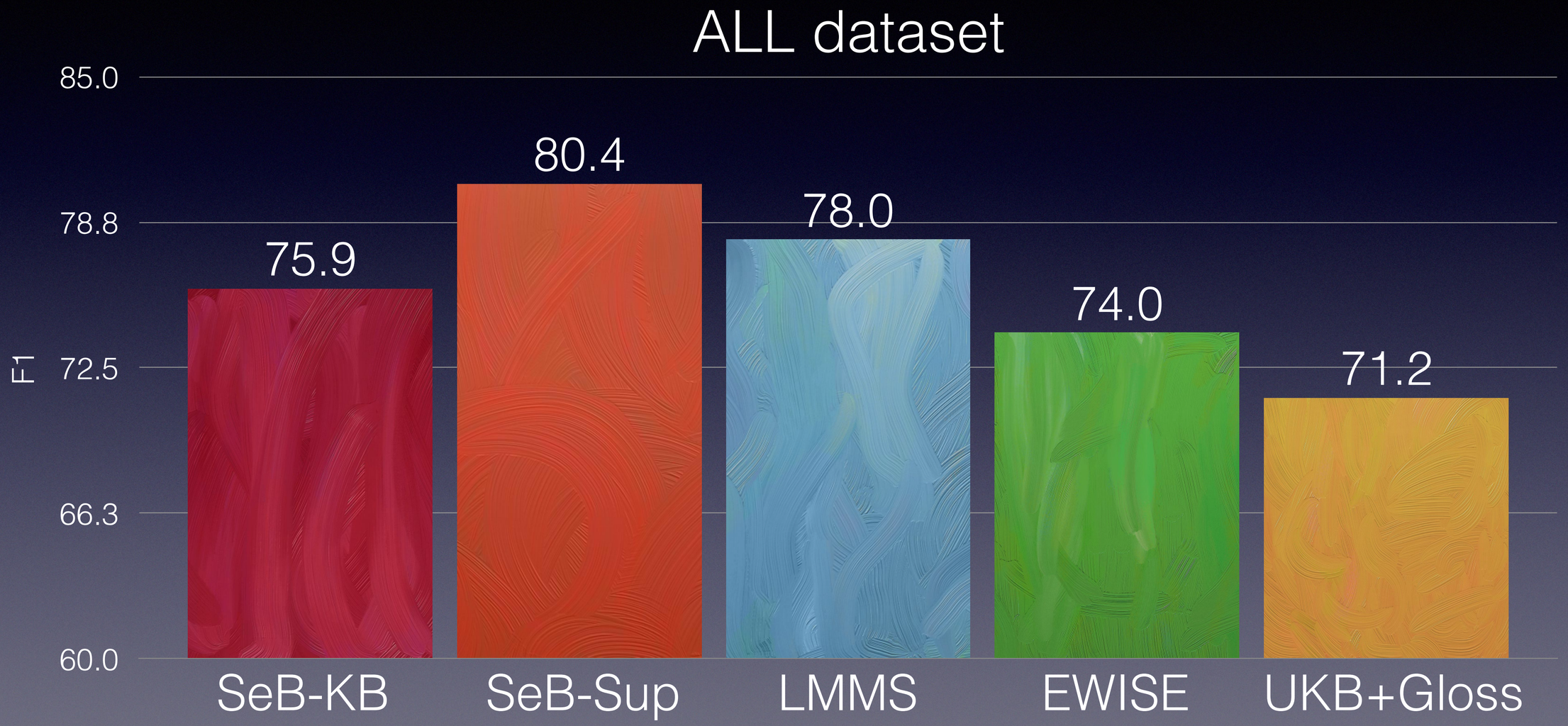
We used **Bert-large-cased for English** and **Bert-base-multilingual-cased** for all the **other languages**.

English Experimental Setup



- **Task:** English Word Sense Disambiguation.
- **Datasets:** Nominal instances of Senseval-2, Senseval-3, SemEval-2007, SemEval-2013, SemEval-2015 and ALL
- **Measure:** F1.
- **Comparison Models:** LMMS (Loureiro and Jorge 2019) and the best performing supervised (EWISE - Kumar et al. 2019) and knowledge-based (UKB - Agirre et al. 2014,2018) approaches.

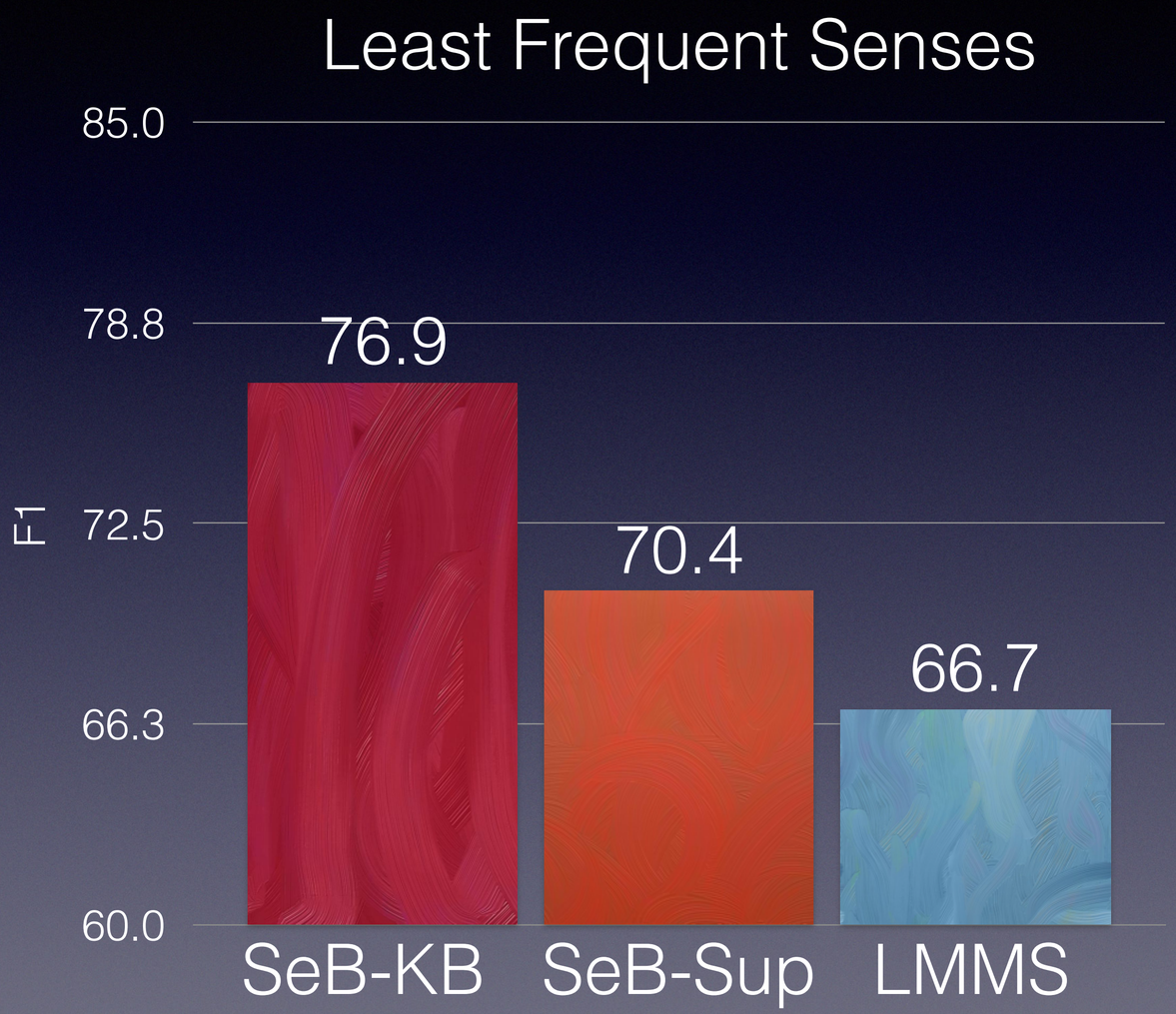
English Results

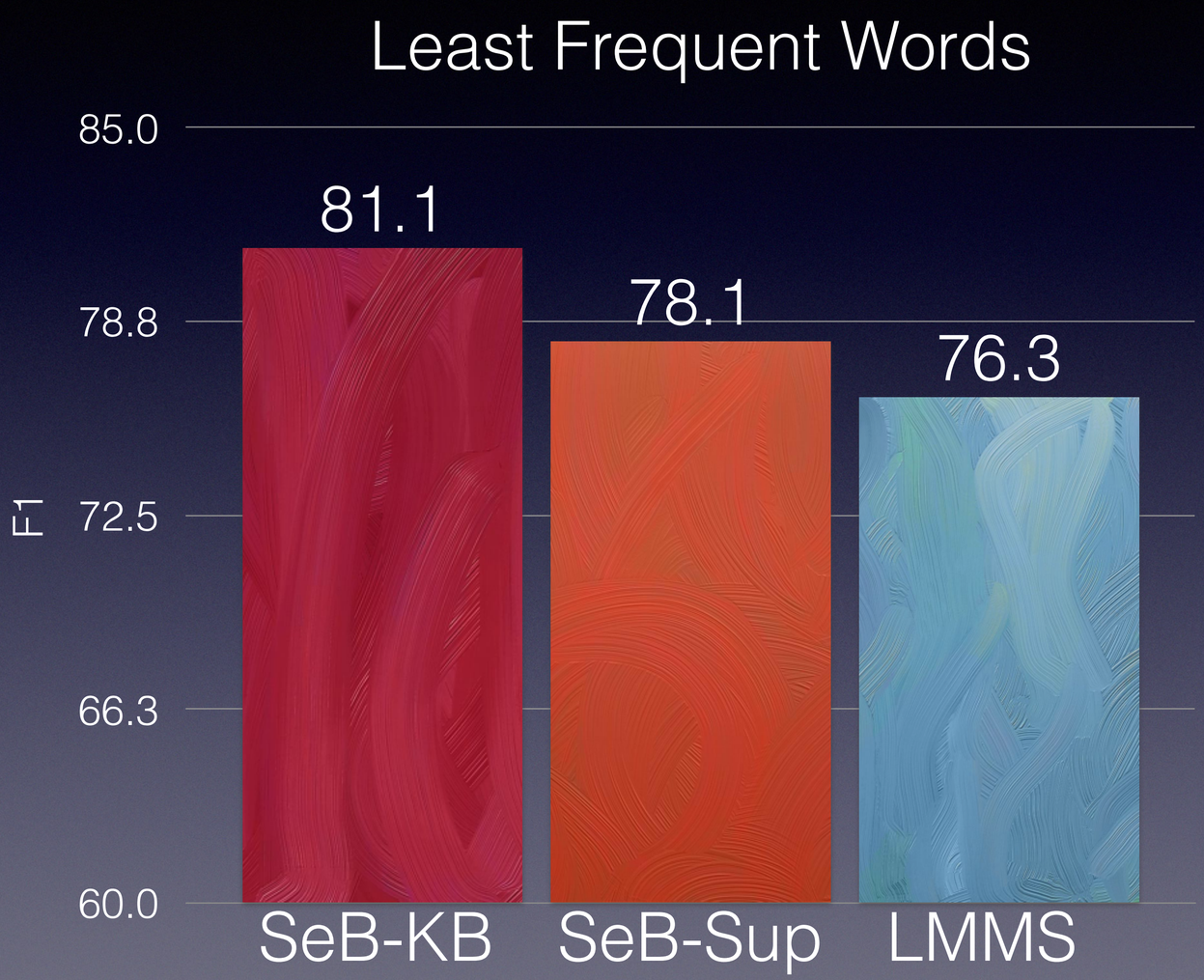


Rare Instances Experimental Setup

- **Task:** English Word Sense Disambiguation.
- **Datasets:** Rare words and senses of the ALL dataset.
- **Measure:** F1.
- **Comparison Model:** LMMS (Loureiro and Jorge 2019).

Rare Instances Results





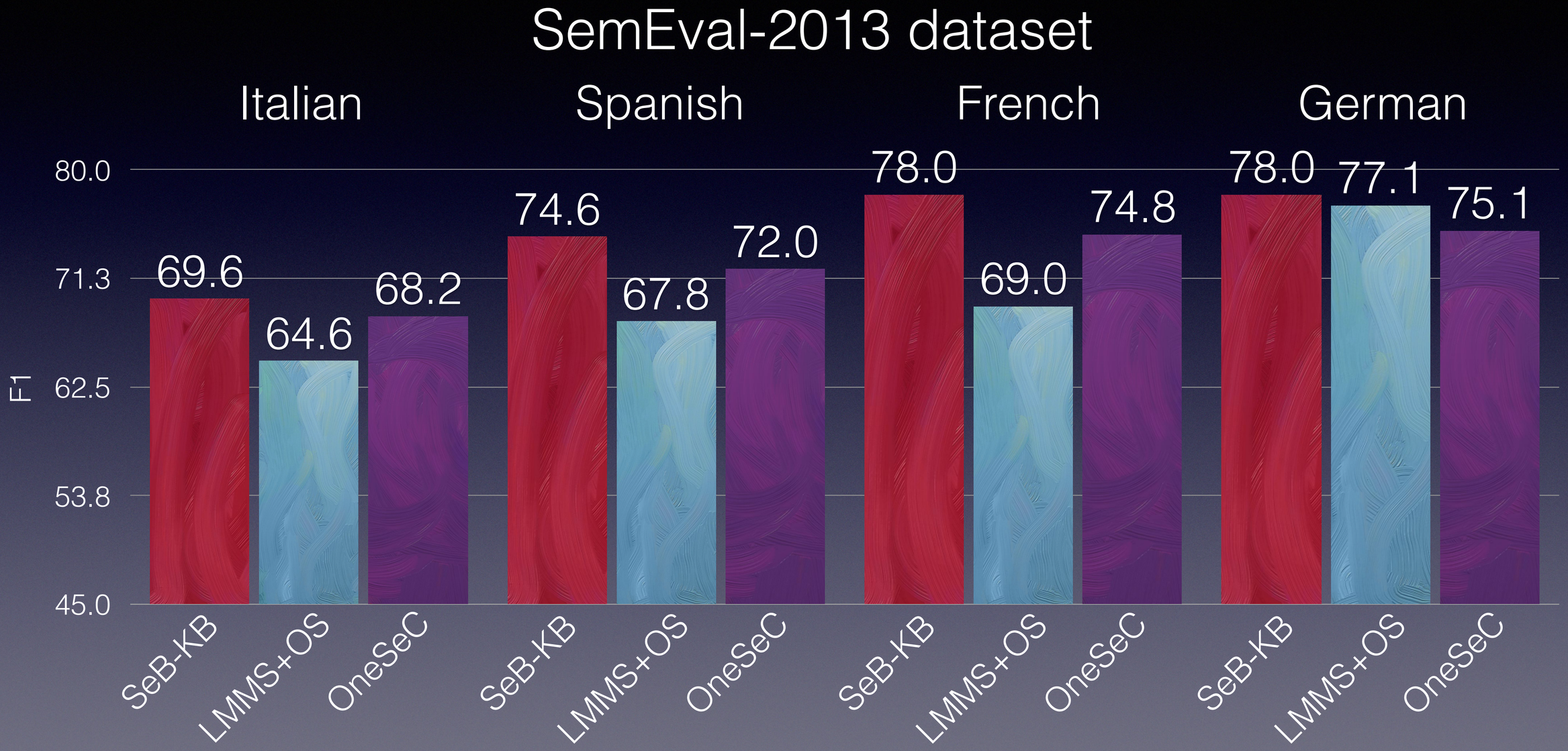
SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

60

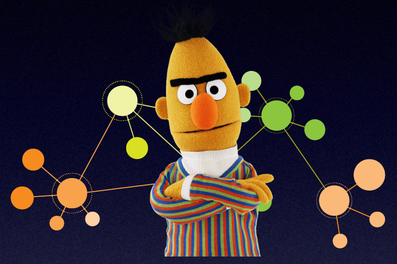
Multilingual Experimental Setup

- **Task:** Multilingual Word Sense Disambiguation in Italian, Spanish, French and German.
- **Datasets:** Italian, Spanish, French and German datasets of SemEval-2013 task 12 and Italian and Spanish datasets of SemEval-2015 task 13.
- **Measure:** F1.
- **Comparison Model:**
  - OneSeC (Scarlini et al. ACL 2019): BiLSTM model trained on the automatically-created datasets for the four languages.
  - LMMS+OneSeC: LMMS vectors trained on OneSeC corpora.

Multilingual Results



Conclusion



We presented SensEmBERT, a multilingual knowledge-based approach for producing embeddings of senses.

SensEmBERT attains SOTA results on the English and multilingual WSD tasks.

We plan to further extend our approach to cover senses with other POS tags.

Vectors Available at: sensembert.org

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

Come to visit our **Poster**

Tonight

6:30 – 8:30 PM

NLP track

SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

Thanks for your attention.

Questions?

sensembert.org

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

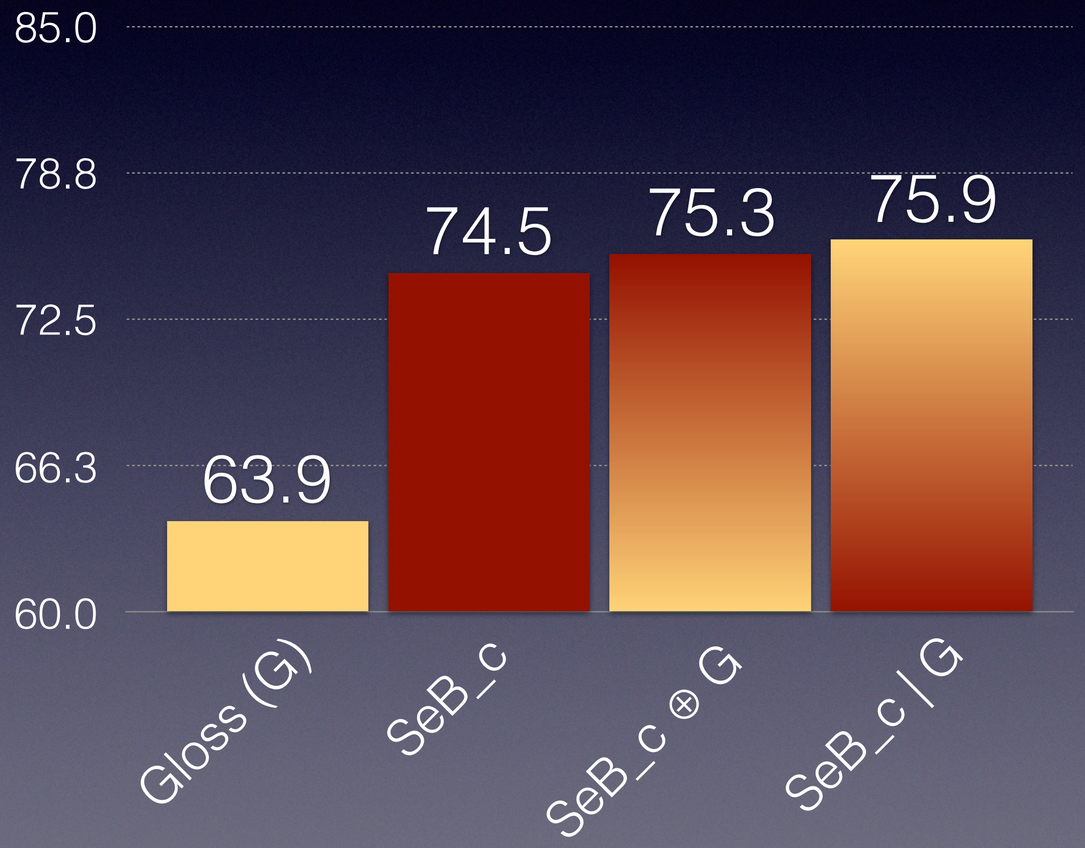
SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual WSD
Bianca Scarlini, Tommaso Pasini and Roberto Navigli

Ablation Study

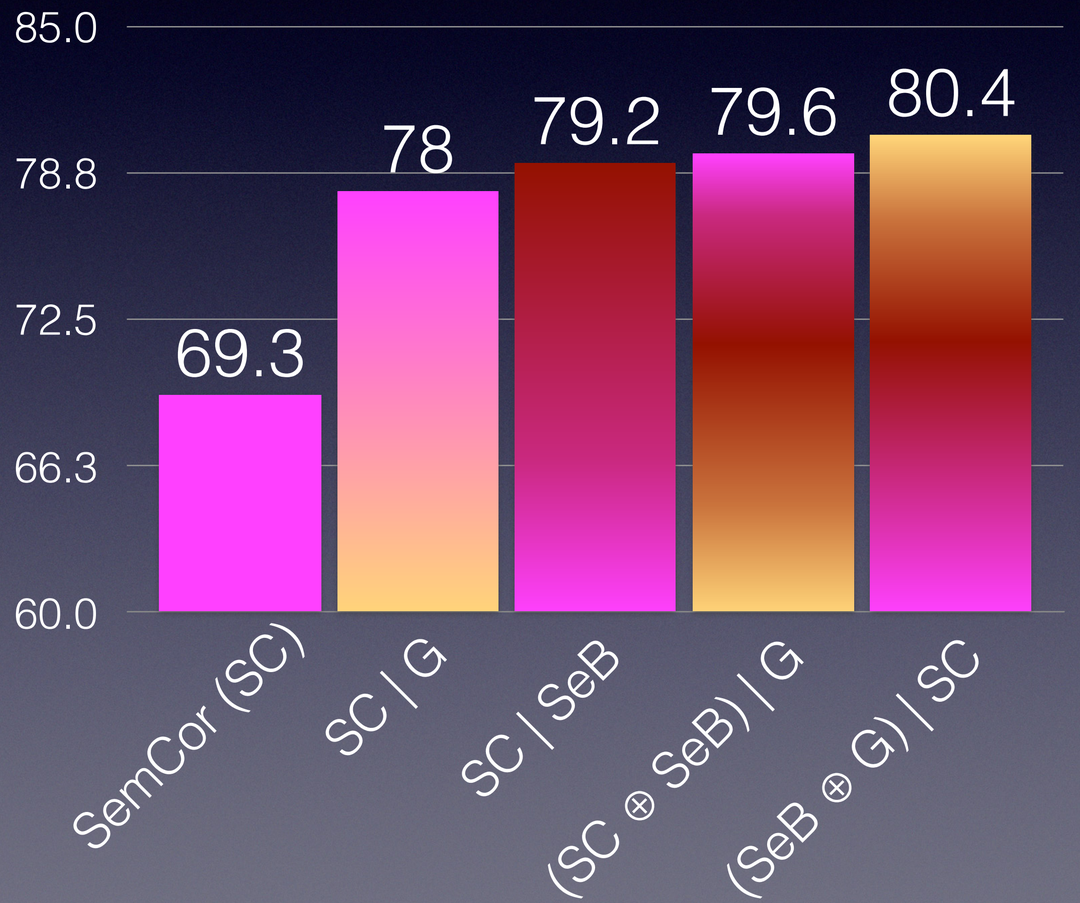
- **Task:** English Word Sense Disambiguation.
- **Datasets:** ALL, i.e., the concatenation of all the nominal instances in the datasets of the evaluation Framework introduced by Raganato et al. 2017.
- **Measure:** F1.

Results

Knowledge Based



Supervised



English Results

	Model	Senseval-2	Senseval-3	SemEval-07	SemEval-13	SemEval-15	ALL
KB	MFS	72.1	72.0	65.4	63.0	66.3	67.6
	Lesk$_{ext}$+emb (2014)	74.6	72.7	66.0	66.2	67.8	69.8
	UKB$_{gloss}$ (2014)	70.6	58.4	56.6	59.0	62.3	62.1
	Babelfy (2014)	74.0	66.7	61.0	66.4	69.9	68.6
Sup	IMS+emb (2016)	79.0	74.6	71.1	65.9	72.1	71.9
	Bi-LSTM (2017)	78.6	72.7	71.1	66.4	73.3	71.6
	HCAN (2018)	78.3	73.2	70.9	68.5	73.8	72.6
	EWISE$_{ConvE}$ (2019)	-	-	-	69.4	-	74.0
Sup$_{context}$	context2vec (2016)	78.0	73.1	66.7	65.6	71.6	71.0
	LSTM-LP (2016)	79.6	76.3	71.7	69.5	72.8	-
	BERT k-NN (2019)	71.7	73.0	72.9	65.6	68.4	69.3
	BERT k-NN + MFS (2019)	81.4	76.3	73.6	71.8	74.0	75.5
	LMMS (2019)	81.7	78.7	78.0	75.1	78.2	78.0
Ours	SENSEMBERT	80.6	70.3	73.6	74.8	**80.2**	75.9
	SENSEMBERT$_{sup}$	**83.7**	**79.7**	**79.9**	**78.7**	**80.2**	**80.4**

Multilingual Results

Model	SemEval-13				SemEval-15	
	IT	ES	FR	DE	IT	ES
Bi-LSTM$_{OneSeC}$	68.2	72.0	74.8	75.1	62.5	62.8
LMMS$_{OneSeC}$	64.6	67.8	69.0	77.1	62.8	49.4
SENSEMBERT	**69.6**	**74.6**	**78.0**	**78.0**	**66.0**	**64.1**