Evaluation of Computational methods for Lexical Semantic Change Detection

Pierluigi Cassotti, pierluigi.cassotti@gu.se

Hello!

Pierluigi Cassotti

Researcher @ Change is Key!
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg (Sweden)
pierluigi.cassotti@gu.se

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling

- Human annotation of Lexical Semantic Change

- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - Generative Models

- Hands-on

Outline

- Computational Modeling Lexical Semantics
  - **Synchronic Modeling**
  - Diachronic Modeling

- Human annotation of Lexical Semantic Change

- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - Generative Models

- Hands-on

Computational Modeling
Lexical Semantics

Word-in-Context (WiC)

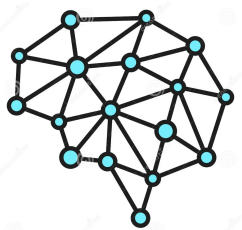
*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

Word-in-Context (WiC)

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

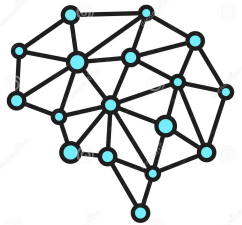
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



Word-in-Context (WiC)

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

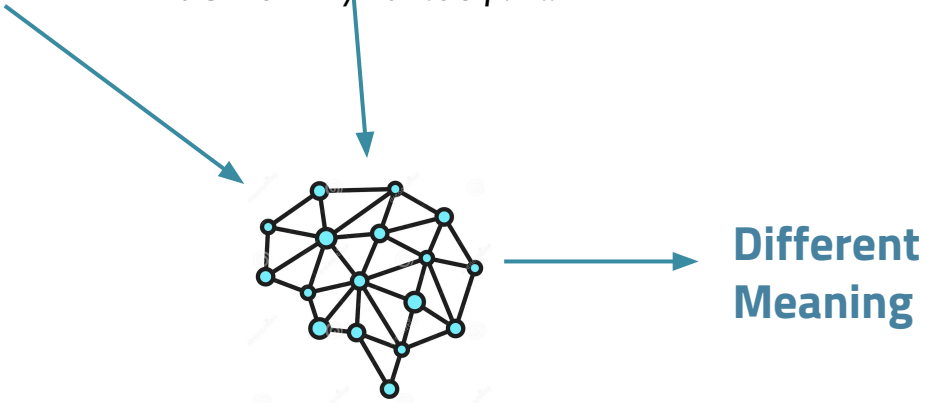
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



Word-in-Context (WiC)

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

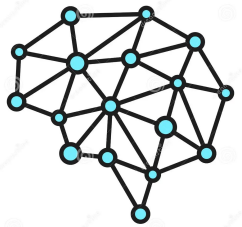
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



Word Sense Disambiguation (WSD)

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

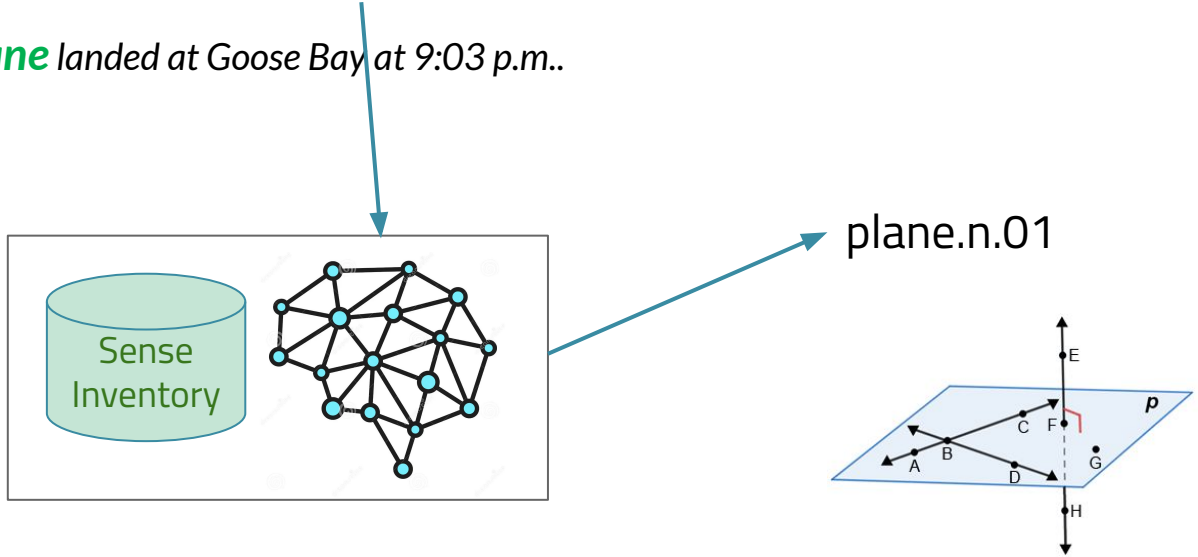
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



Word Sense Disambiguation (WSD)

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

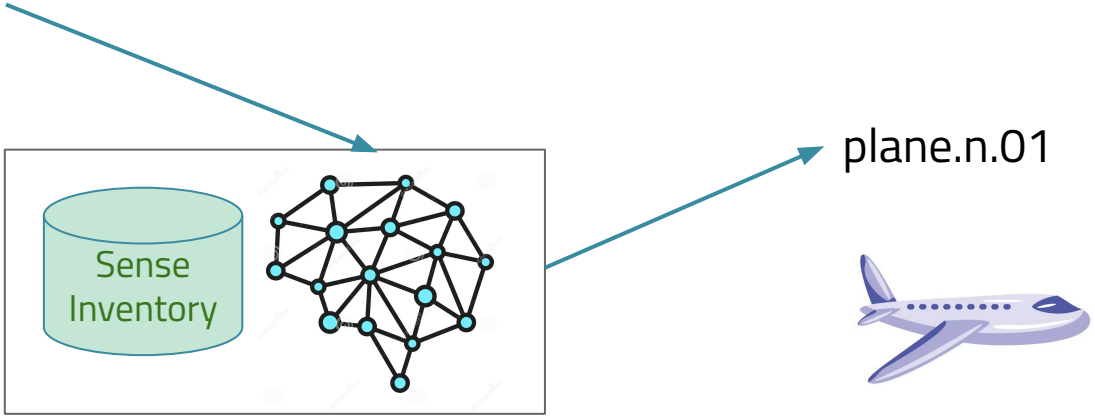
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



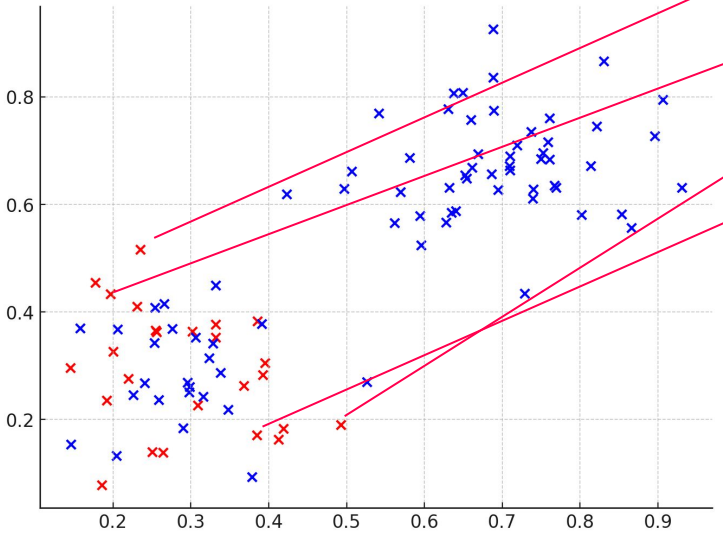
Word Sense Disambiguation (WSD)

Provide a large table; this is a horizontal ***plane****, and will represent the ground plane, viz.*

The President's ***plane*** *landed at Goose Bay at 9:03 p.m..*



Word Sense Induction (WSI)



Representation

If the sun's rays be parallel to any **plane**, that plane to which they are parallel, is called a plane of shade.

its angles upon a given point A, in the **plane**, on which the ichnography is to be described;

will be no difference between the shadow on the **plane**, and the side of the prism which projects that shadow;

There are other kinds of **planes** besides the above; as the plough, for sinking a groove to receive a projecting tongue;

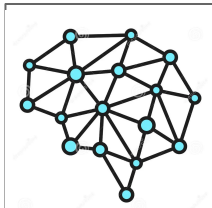
Troy turned it to the right, and the **plane** turned to the right, just

They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the **planes**;

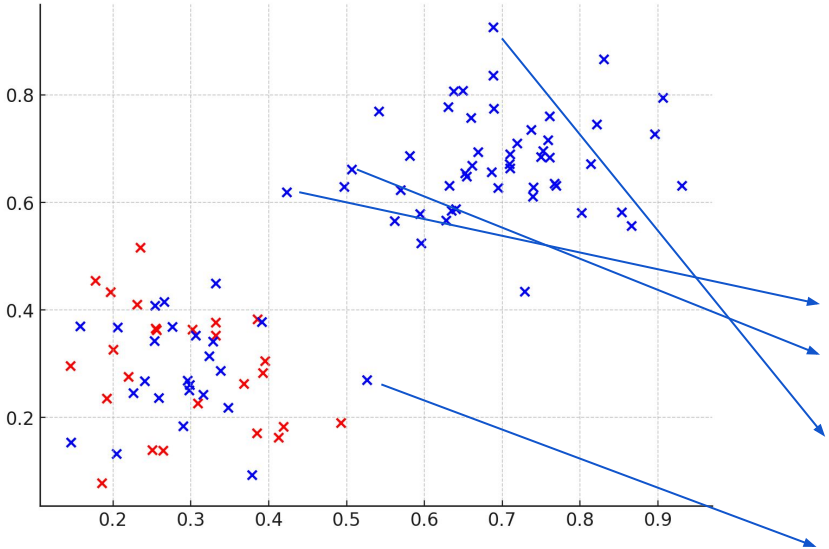
In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body **planes** leased from commercial airlines.

Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal **plane** (Rab &; Grottkau, 2001).

Word Sense Induction (WSI)



Representation



If the sun's rays be parallel to any **plane**, that plane to which they are parallel, is called a plane of shade.

its angles upon a given point A, in the **plane**, on which the ichnography is to be described;

will be no difference between the shadow on the **plane**, and the side of the prism which projects that shadow;

There are other kinds of **planes** besides the above; as the plough, for sinking a groove to receive a projecting tongue;

Troy turned it to the right, and the **plane** turned to the right, just

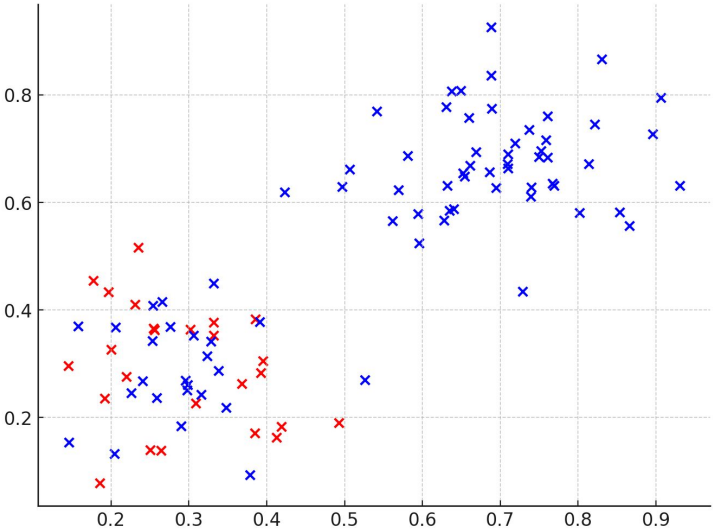
They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the **planes**;

In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body **planes** leased from commercial airlines.

Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal **plane** (Rab &; Grottkau, 2001).

Word Sense Induction (WSI)

Sense Induction



If the sun's rays be parallel to any **plane**, that plane to which they are parallel, is called a plane of shade.

its angles upon a given point A, in the **plane**, on which the ichnography is to be described;

will be no difference between the shadow on the **plane**, and the side of the prism which projects that shadow;

There are other kinds of **planes** besides the above; as the plough, for sinking a groove to receive a projecting tongue;

Troy turned it to the right, and the **plane** turned to the right, just

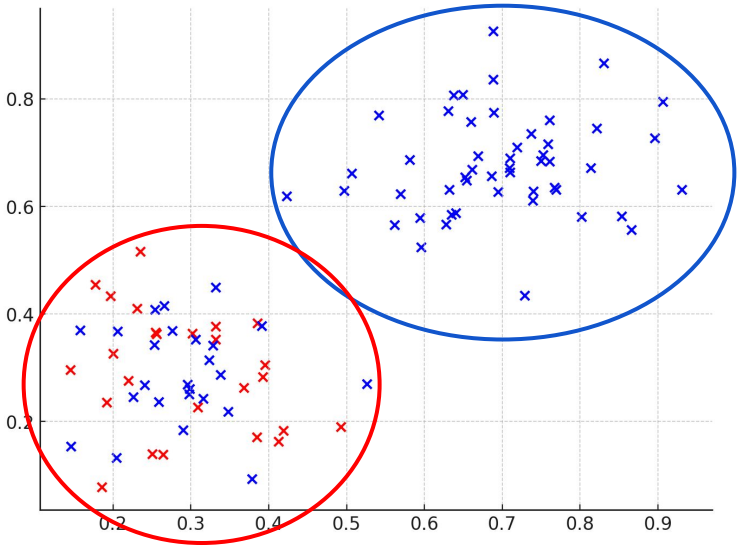
They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the **planes**;

In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body **planes** leased from commercial airlines.

Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal **plane** (Rab &; Grottkau, 2001).

Word Sense Induction (WSI)

Sense Induction



If the sun's rays be parallel to any **plane**, that plane to which they are parallel, is called a plane of shade.

its angles upon a given point A, in the **plane**, on which the ichnography is to be described;

will be no difference between the shadow on the **plane**, and the side of the prism which projects that shadow;

There are other kinds of **planes** besides the above; as the plough, for sinking a groove to receive a projecting tongue;

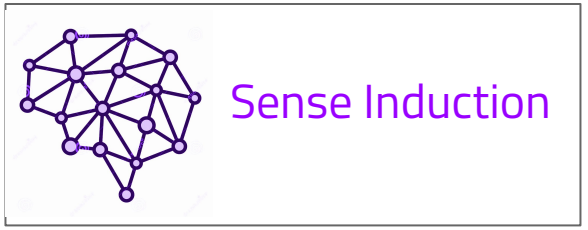
Troy turned it to the right, and the **plane** turned to the right, just

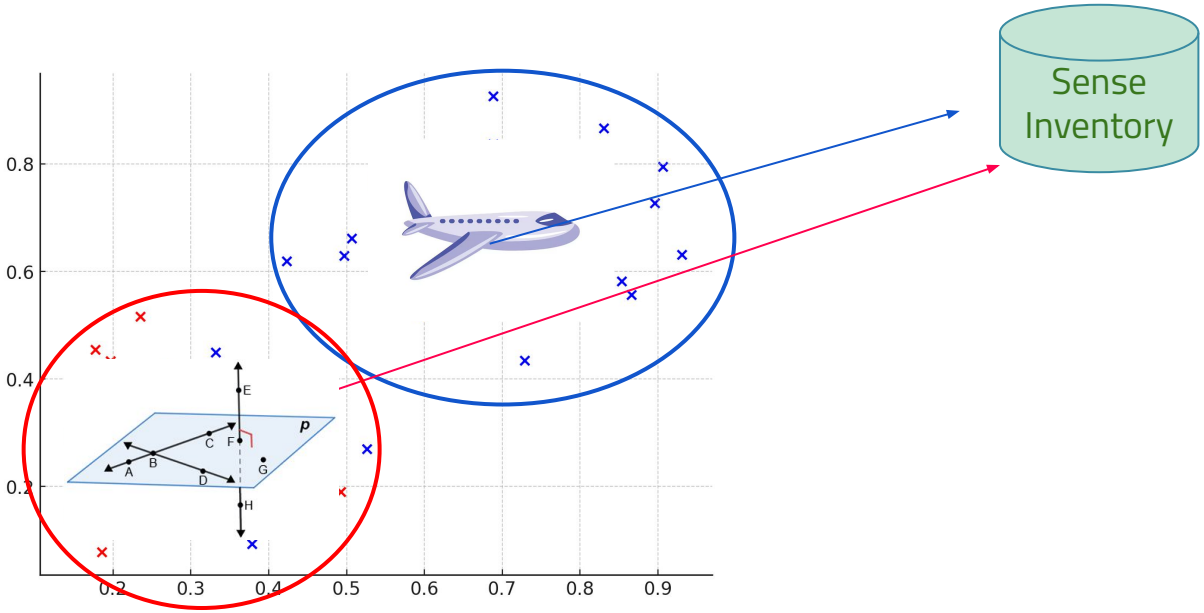
They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the **planes**;

In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body **planes** leased from commercial airlines.

Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal **plane** (Rab &; Grottkau, 2001).

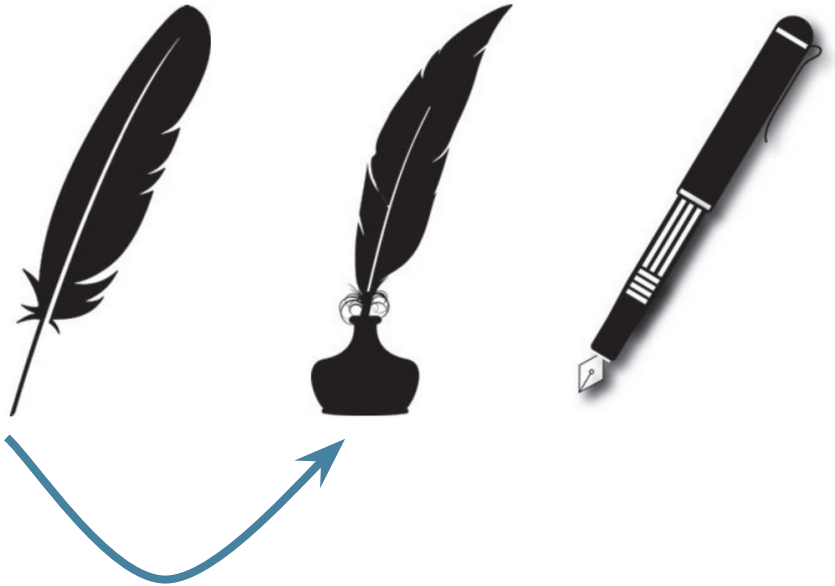
Word Sense Induction (WSI)



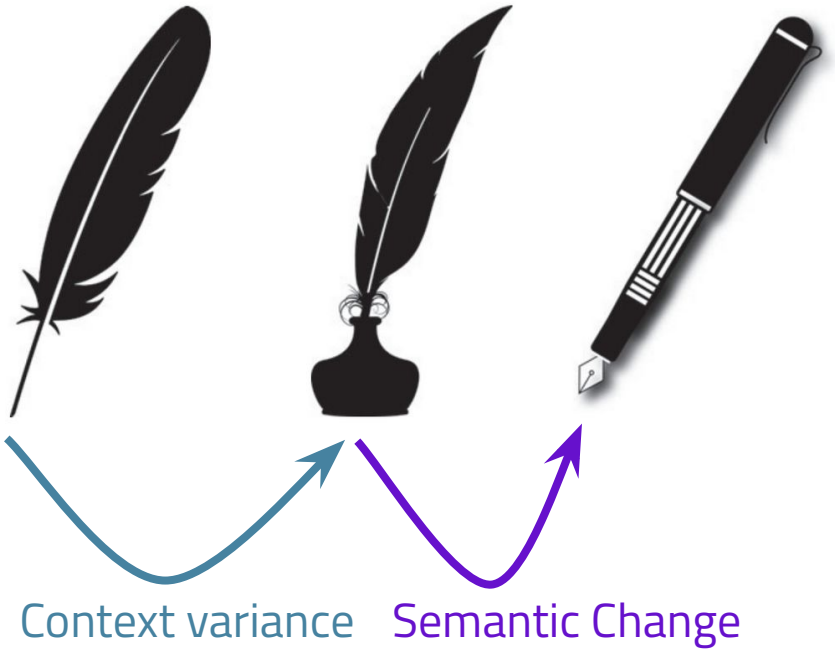


Semasiological change

Semasiological change



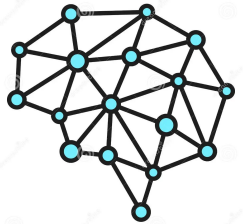
Semasiological change



Definition Modeling

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

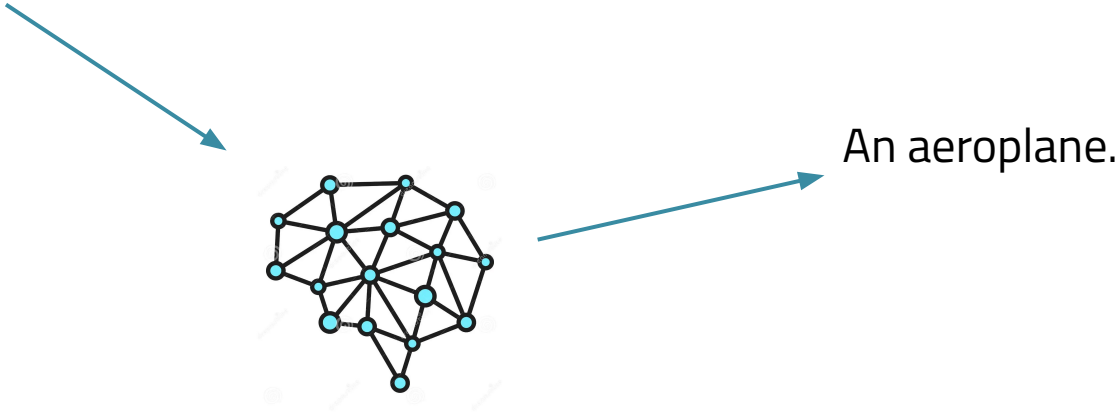


A flat or level surface of a material body

Definition Modeling

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

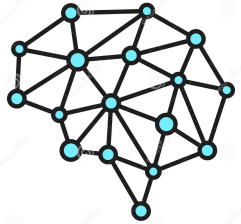
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



Lexical Substitution

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

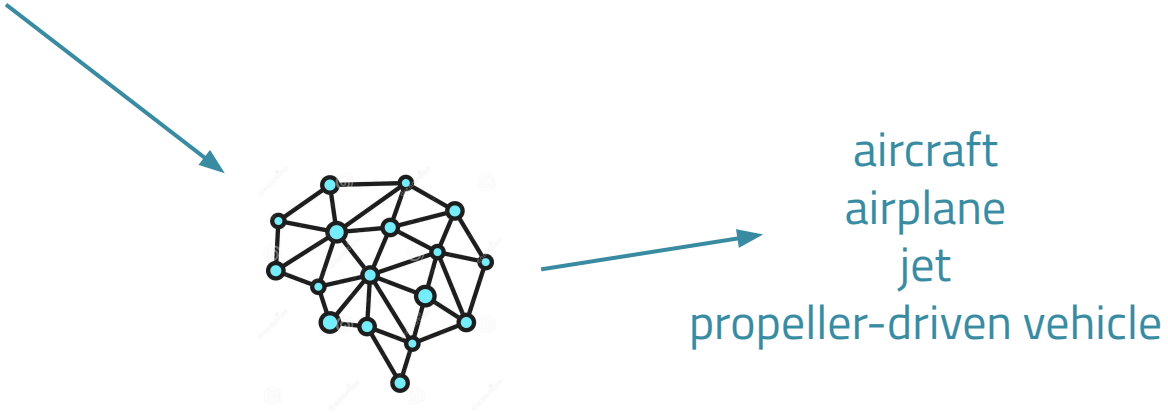


level
surface
flat plane
horizontal plane

Lexical Substitution

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

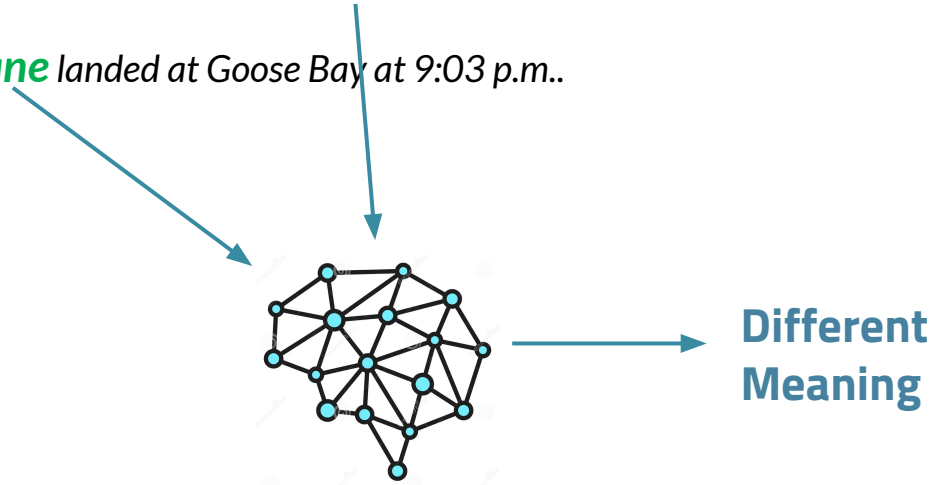


Approach properties: Sense Discrimination

Word-in-Context (WiC) do **NOT explicit** assign sense labels

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

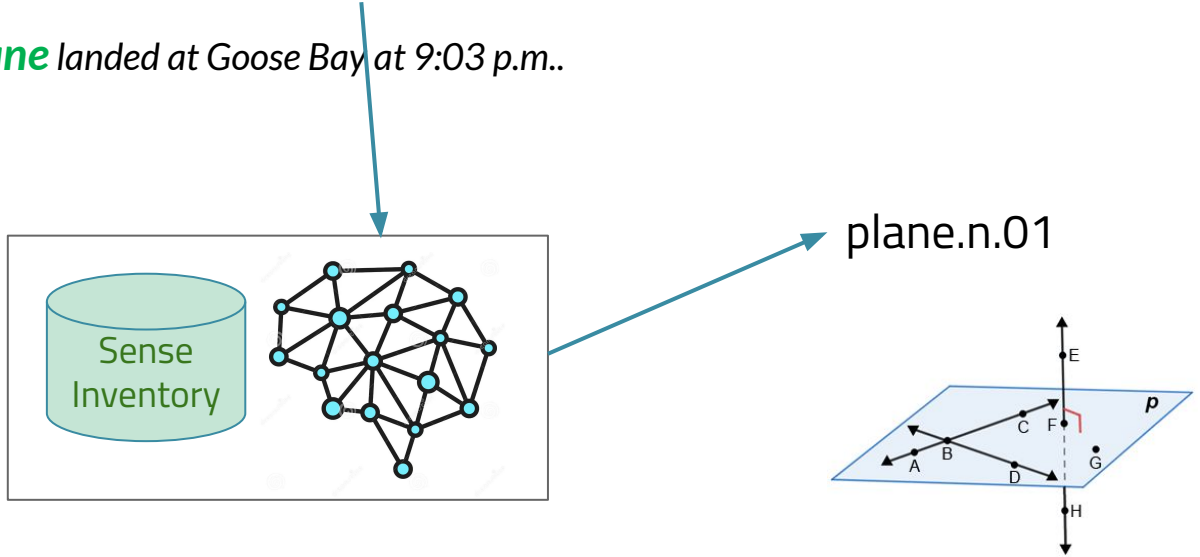


Approach properties: Sense Discrimination

Word Sense Disambiguation (WSD) explicitly assign sense labels

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

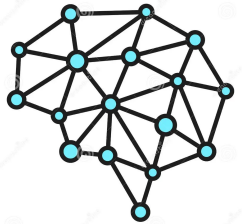


Approach properties: Interpretability

Word-in-Context (WiC) is **NOT interpretable**

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*



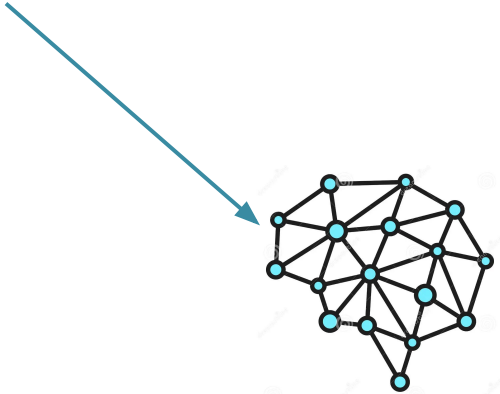
Different Meaning

Approach properties: Interpretability

Lexical substitutions can be used to **interpret the results**

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*



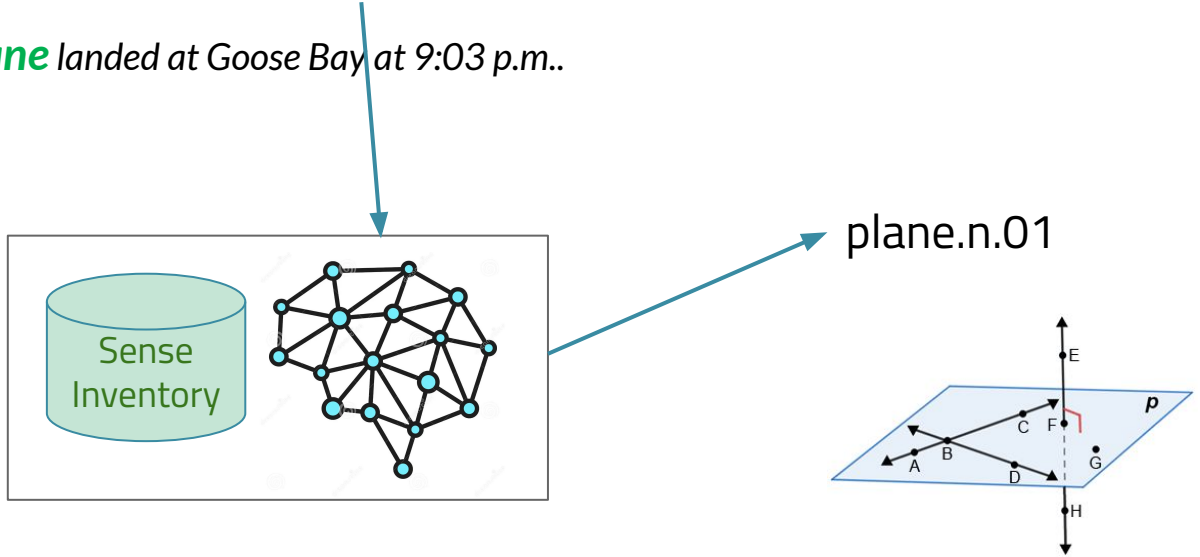
aircraft
airplane
jet
propeller-driven vehicle

Approach properties: Sense Inventory Independent

Word Sense Disambiguation (WSD) rely on a **specific** sense inventory

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

*The President's **plane** landed at Goose Bay at 9:03 p.m..*

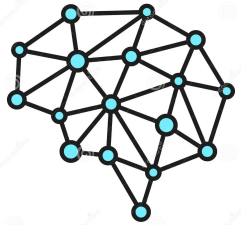


Approach properties: Sense Inventory Independent

Definition Modeling
generate sense definitions that are sense inventory **independent**

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

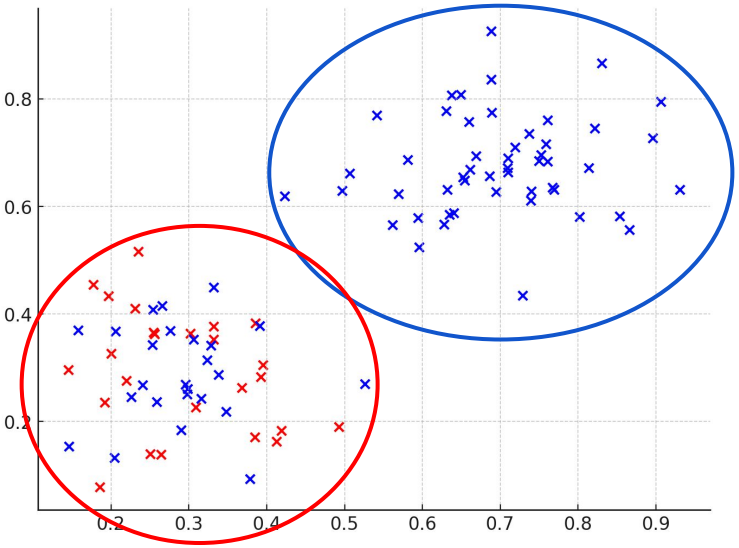
*The President's **plane** landed at Goose Bay at 9:03 p.m..*



A flat or level surface of a material body

Approach properties: Parameter free

WSI is **sensitive** to parameters tuning in the sense induction phase



If the sun's rays be parallel to any **plane**, that plane to which they are parallel, is called a plane of shade.

its angles upon a given point A, in the **plane**, on which the ichnography is to be described;

will be no difference between the shadow on the **plane**, and the side of the prism which projects that shadow;

There are other kinds of **planes** besides the above; as the plough, for sinking a groove to receive a projecting tongue;

Troy turned it to the right, and the **plane** turned to the right, just

They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the **planes**;

In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body **planes** leased from commercial airlines.

Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal **plane** (Rab &; Grottkau, 2001).

Summary

	Sense Differentiation	*Interpretability*	*Sense Inventory independent*	*Parameter free*	*Explicit Granularity tuning*	*Analysis Level*
Word-in-Context (WiC)	NO	NO	YES	YES	NO	Pair
Word Sense Induction (WSI)	YES	YES	YES	NO	NO	Corpus
Word Sense Disambiguation (WSD)	YES	YES	NO	YES	YES	Sentence
Target Sense Verification (TSV)	YES	YES	NO	YES	NO	Sentence
Lexical Substitution	NO	YES	YES	YES	NO	Sentence
Definition Modeling	NO	YES	YES	YES	YES	Sentence

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - **Diachronic Modeling**

- Human annotation of Lexical Semantic Change

- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - Generative Models

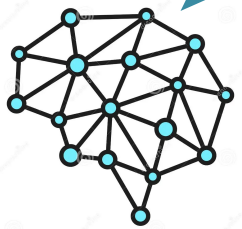
- Hands-on

Semantics in Diachrony

Diachronic Word-in-Context (DWiC)

December 15th 1836 — *Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

June 2th 1999 — *The President's **plane** landed at Goose Bay at 9:03 p.m..*

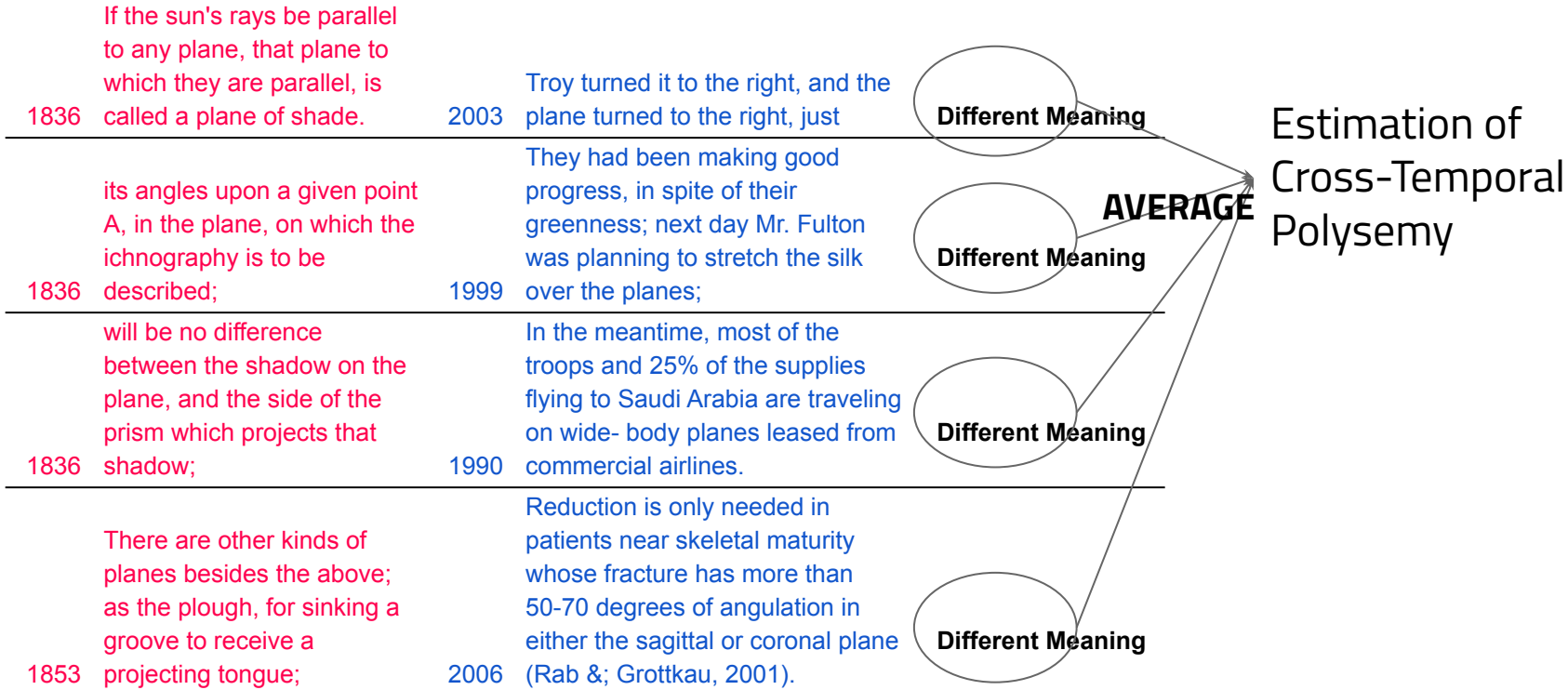


Different Meaning

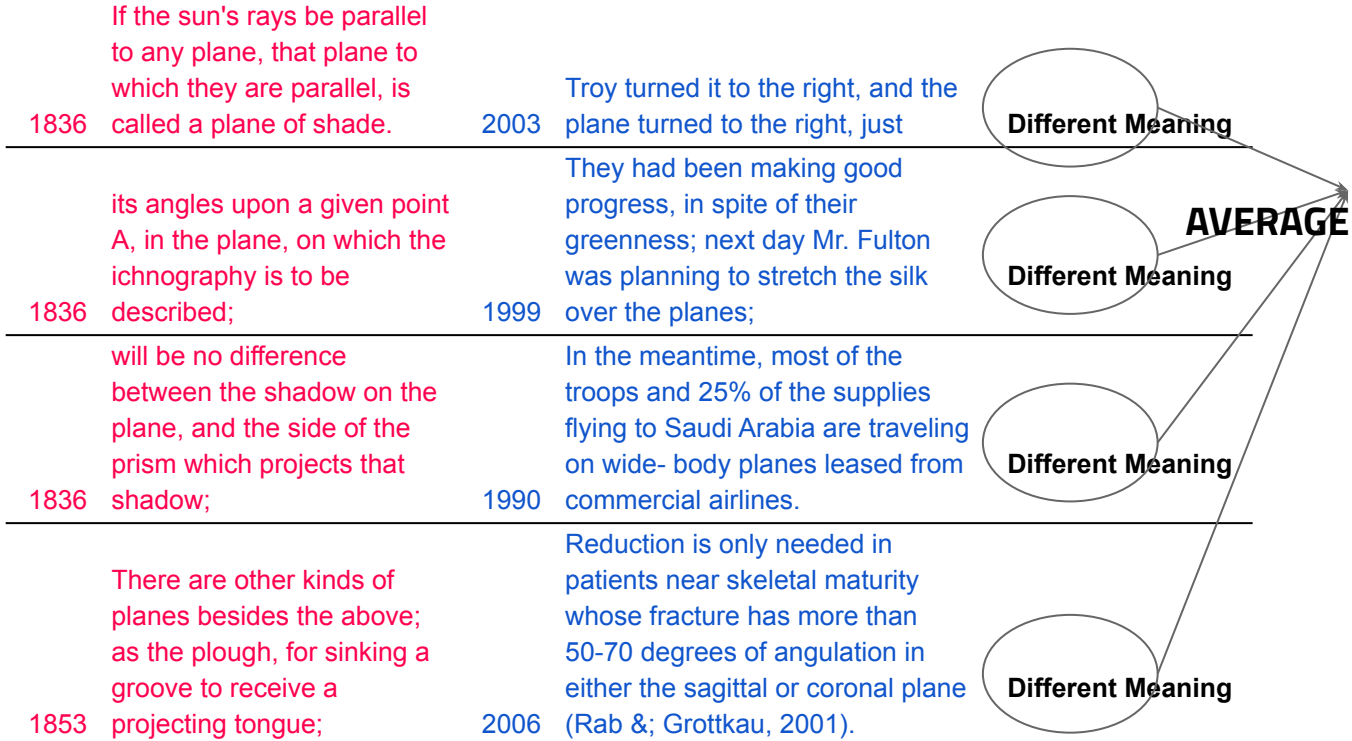
Diachronic Word-in-Context (DWiC)

1836	If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.	2003	Troy turned it to the right, and the plane turned to the right, just	**Different Meaning**
1836	its angles upon a given point A, in the plane, on which the ichnography is to be described;	1999	They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;	**Different Meaning**
1836	will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;	1990	In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.	**Different Meaning**
1853	There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;	2006	Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).	**Different Meaning**

Diachronic Word-in-Context (DWiC)



Diachronic Word-in-Context (DWiC)

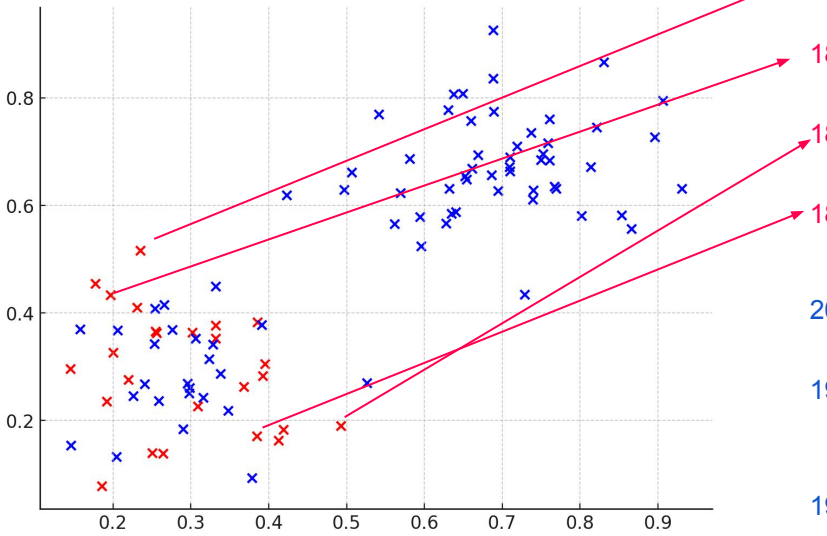


Estimation of Cross-Temporal Polysemy

↓

COMPARE MEASURE (Schlechtweg et al.,2018)

Diachronic Word Sense Induction



1836 If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.

1836 its angles upon a given point A, in the plane, on which the ichnography is to be described;

1836 will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;

1853 There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

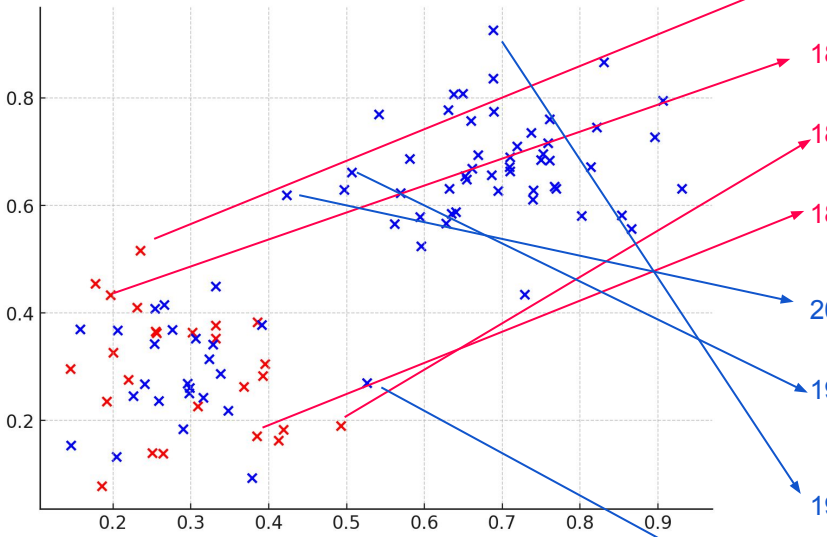
2003 Troy turned it to the right, and the plane turned to the right, just

1999 They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;

1990 In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.

2006 Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Diachronic Word Sense Induction



1836 If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.

1836 its angles upon a given point A, in the plane, on which the ichnography is to be described;

1836 will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;

1853 There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

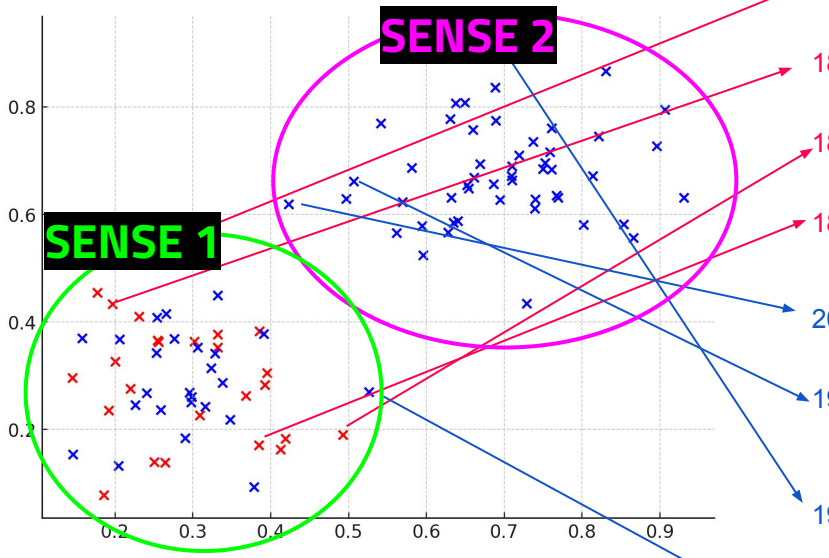
2003 Troy turned it to the right, and the plane turned to the right, just

1999 They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;

1990 In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.

2006 Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Diachronic Word Sense Induction



1836 If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.

1836 its angles upon a given point A, in the plane, on which the ichnography is to be described;

1836 will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;

1853 There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

2003 Troy turned it to the right, and the plane turned to the right, just

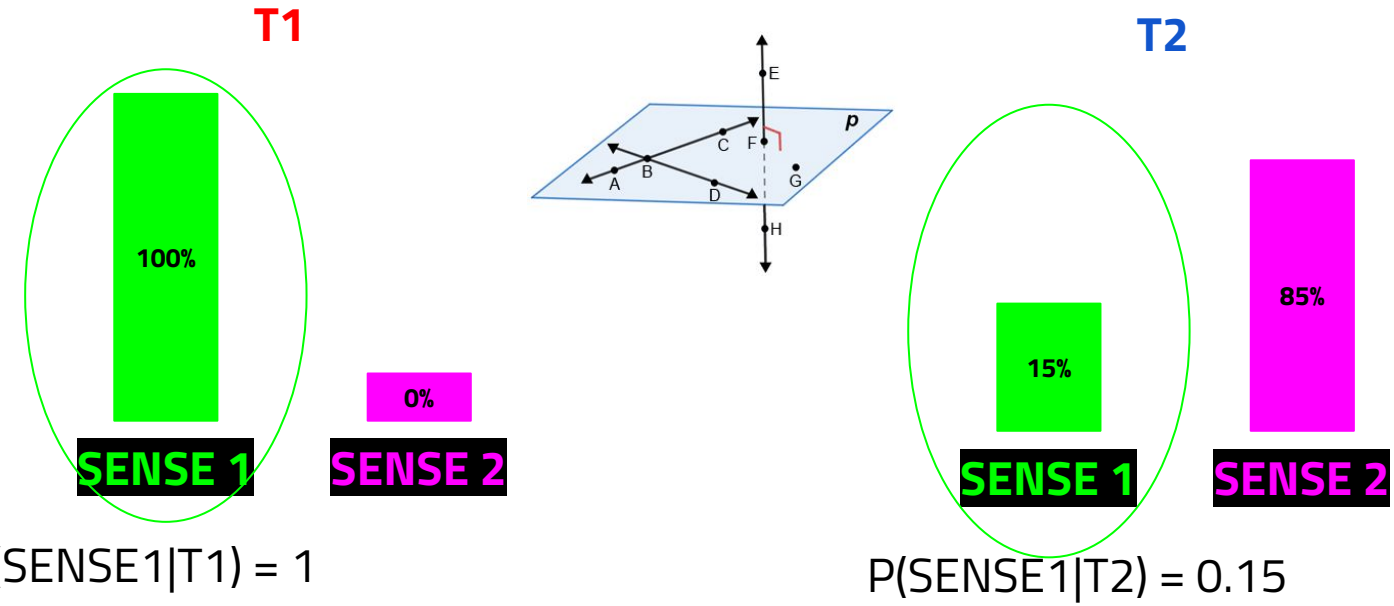
1999 They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;

1990 In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.

2006 Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Diachronic Word Sense Induction

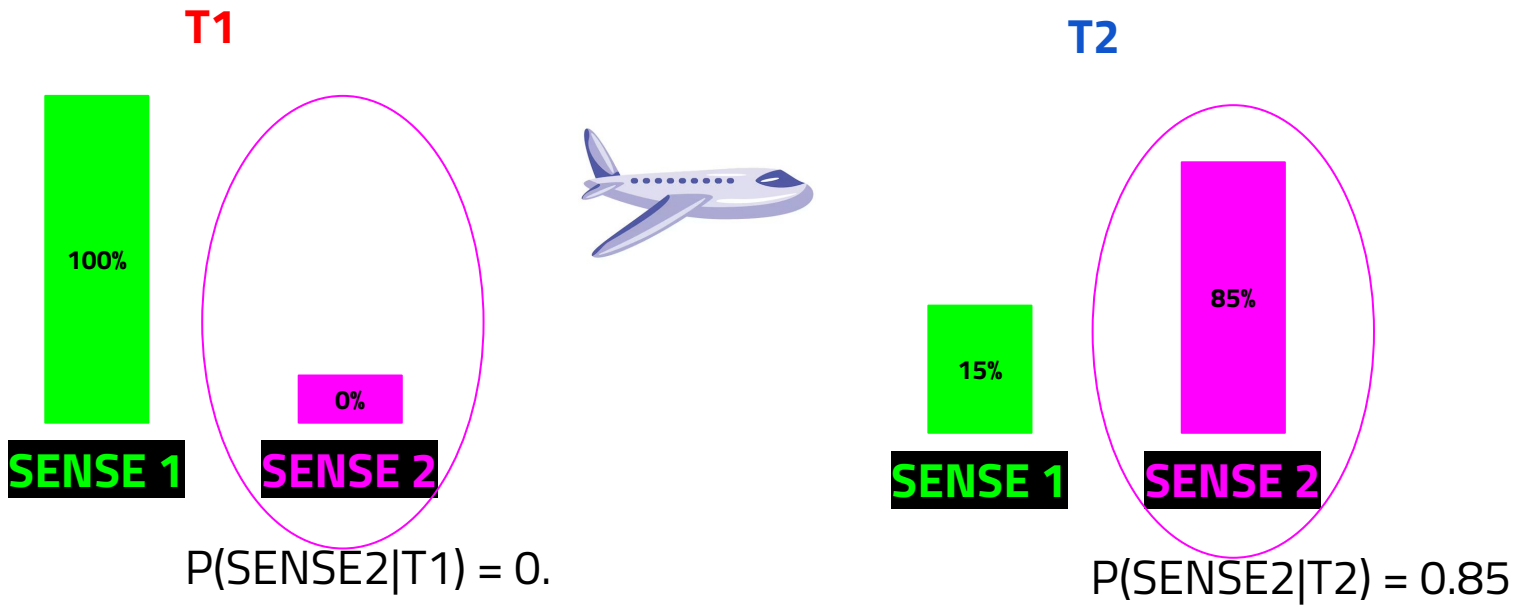
Diachronic Word Sense Induction



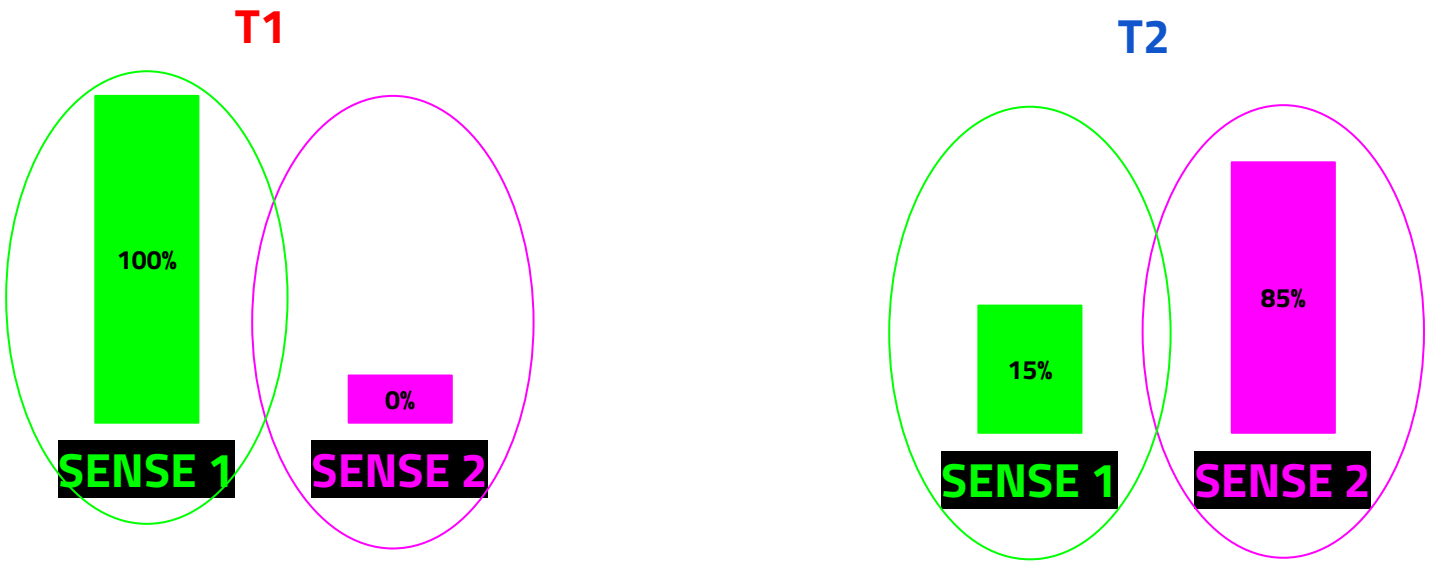
P(SENSE1|T1) = 1

P(SENSE1|T2) = 0.15

Diachronic Word Sense Induction

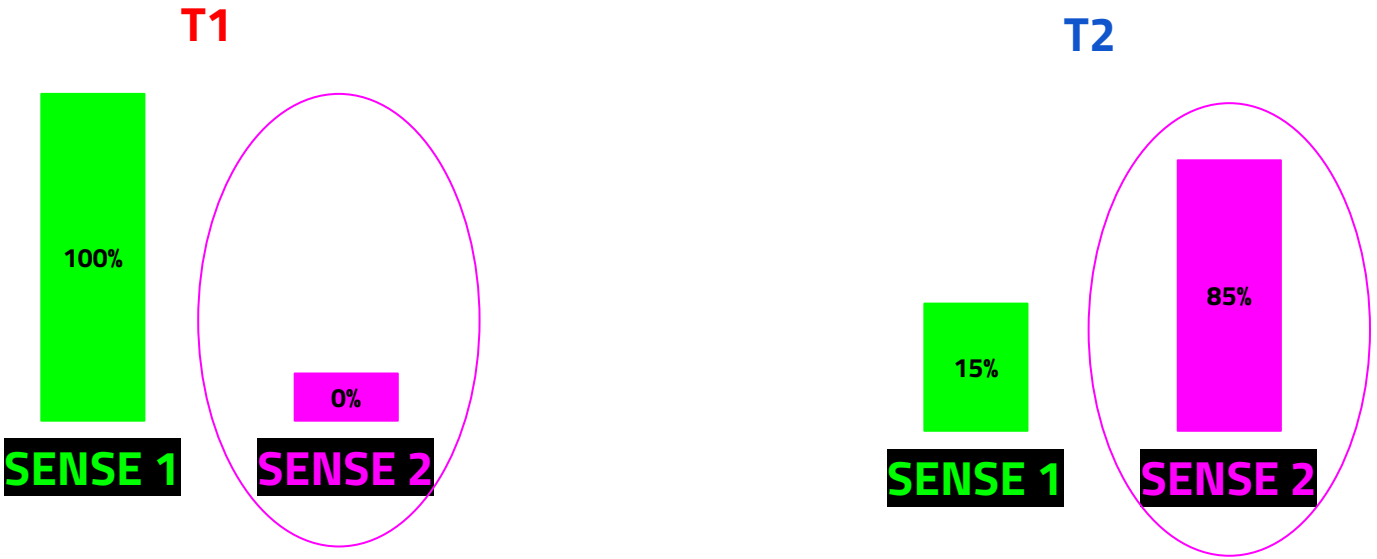


Graded Score of Semantic Change



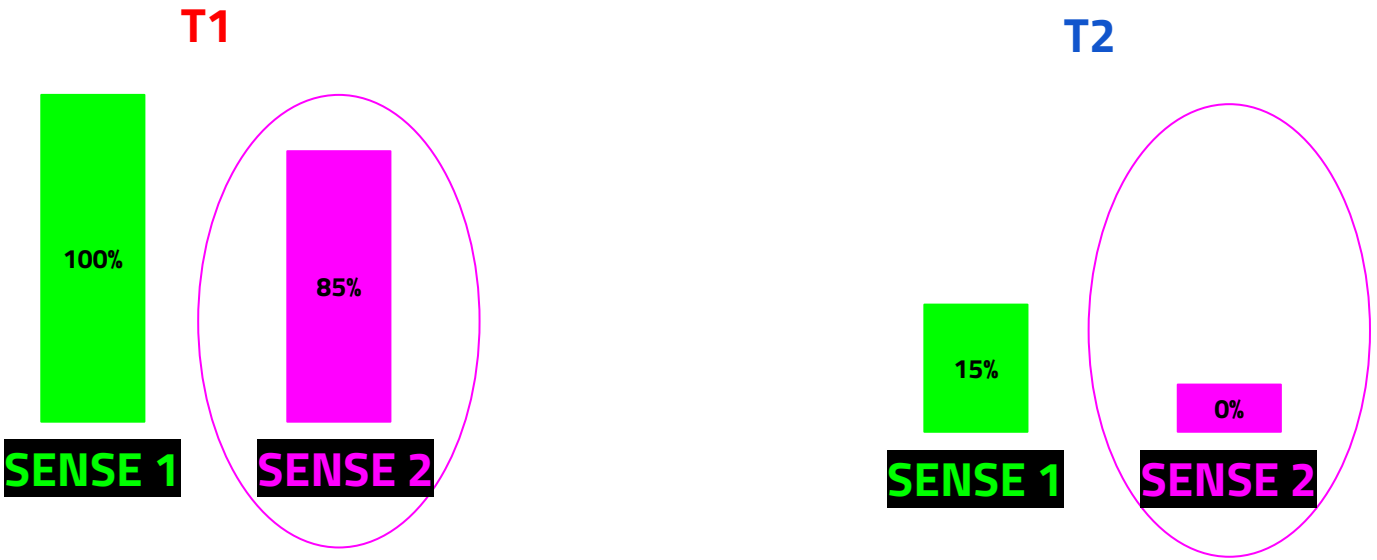
Distance between probability distributions

Binary Label of Semantic Change



If the sense probability < of the threshold in T1 and > of the threshold in T2 the sense is gained

Binary Label of Semantic Change



If the sense probability > of the threshold in T1 and < of the threshold in T2 the sense is lost

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling

- **Human annotation of Lexical Semantic Change**

- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - Generative Models

- Hands-on

DURel

Diachronic Usage Relatedness (DURel):
A Framework for the Annotation of Lexical Semantic Change

*Dominik Schlechtweg, *Sabine Schulte im Walde, †Stefanie Eckmann
*Institute for Natural Language Processing, University of Stuttgart, Germany
†Historical and Indo-European Linguistics, LMU Munich, Germany
dominik.schlechtweg@ims.uni-stuttgart.de, schulte@ims.uni-stuttgart.de,
stefanie.eckmann@campus.lmu.de

Word-in-Context task

For each word W:

1: Identical
0: Unrelated

?

"...she opened a vein in her **arm**..." ⟷ "...overlooking and **arm** of the sea..."

Graded Word-in-Context task

For each word W:

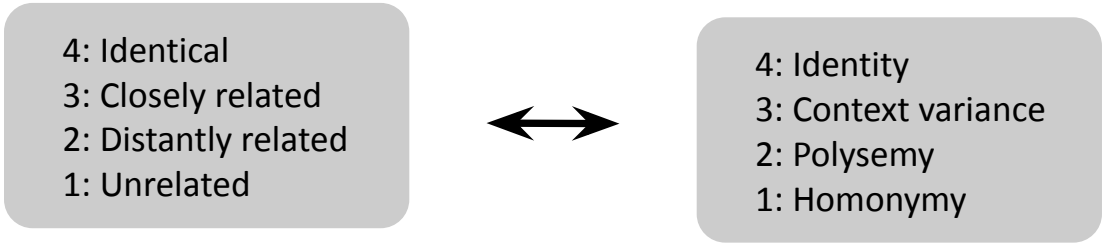
1: Identical
0: Unrelated

?

"...she opened a vein in her **arm**..." ⟷ "...overlooking and **arm** of the sea..."

4: Identical
3: Closely related
2: Distantly related
1: Unrelated

Semantic Relatedness vs Semantic proximity (Blank 1997)



Diachronic Graded Word-in-Context task

For each word W:

?

T1 "...she opened a vein in her **arm**..." ⟷ "...overlooking and **arm** of the sea..." T2

Language	Corpus	T1	T2
English	CCOHA	1810 - 1860	1960 - 2010
German	DTA; BZ+ND	1800 - 1899	1946 - 1990
Latin	LatinISE	-200 - 0	0 - 2000
Swedish	Kubhist	1790 - 1830	1895 - 1903

How did we do it?

For each word W:

"...she opened a vein in her **arm**..."**T1**

"...to pay for the **arms** and ammunition..."**T1**

"...within reach of his **arm**..."**T1**

T2 "...overlooking and **arm** of the sea..."

T2 "...an **arm** of the Red Sea..."

T2 "...when the disembodied **arm** of the Statue..."

How did we do it?

For each word W:

?

"...she opened a vein in her **arm**..."T1

"...to pay for the **arms** and ammunition..."T1

"...within reach of his **arm**..."T1

T2 "...overlooking and **arm** of the sea..."

T2 "...an **arm** of the Red Sea..."

T2 "...when the disembodied **arm** of the Statue..."

How did we do it?

For each word W:

?

"...she opened a vein in her **arm**..."**T1**

"...to pay for the **arms** and ammunition..."**T1**

"...within reach of his **arm**..."**T1**

T2 "...overlooking and **arm** of the sea..."

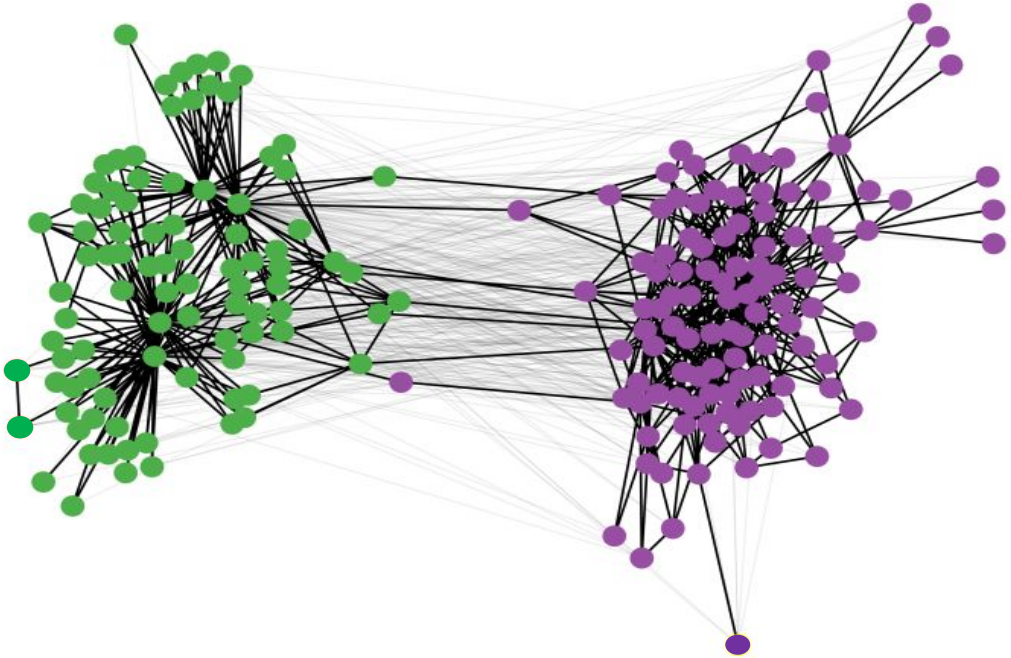
T2 "...an **arm** of the Red Sea..."

T2 "...when the disembodied **arm** of the Statue..."

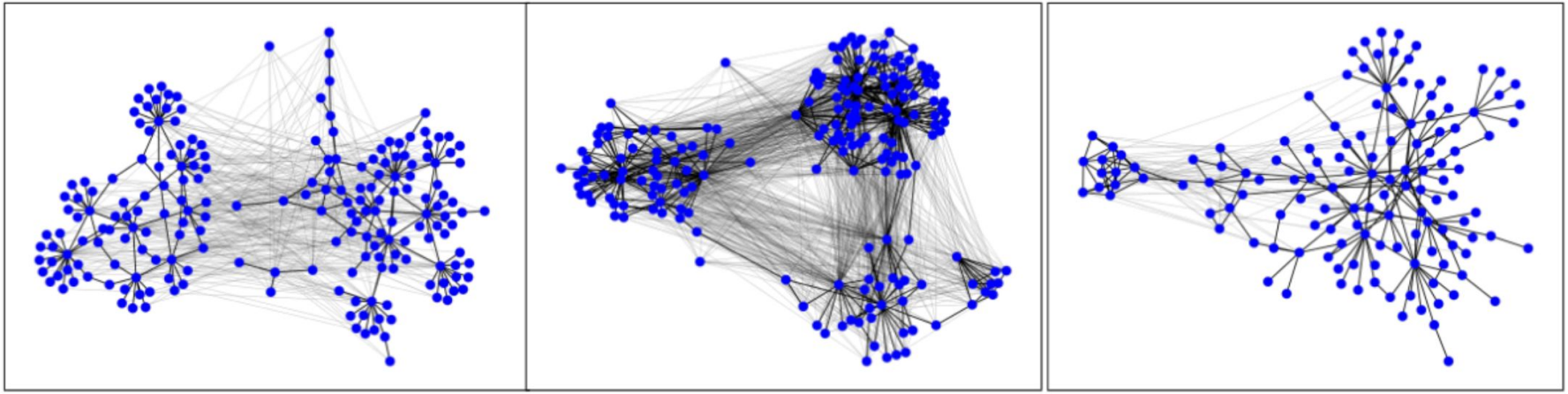
4: Identical
3: Closely related
2: Distantly related
1: Unrelated

T1 Pä Caudia mar fullkomlige » lugnt anda lill flutet af April, dä ön hastigt säg sig blockerad af 31 Cassotista och	4	T1 Pä Caudia mar fullkomlige » lugnt anda lill flutet af April, dä ön hastigt säg sig blockerad af 31 Cassotista och
T1 De äro styfsinta — sade han — men de ge nog med sig, är det kommer till kritan. man fått löfte om hardt när enhälligt understöd från den" nya center	4	T2 dels förtröstade man på, att folkpartiet, när det komme till kritan, hellre skulle taga till och med denna skenreform än — afslag.
T2 kontant efter den i lisra punkten stadgade grund. till det andra genom en oändlig rad gradationer, och alla eiler nästan alla fenomen, tilldragniog, suggestion etc., som man frambringar under sömnen, kunna hos de flesta medier lika val framkallas i vaket tillstånd.	1	T1 löpande årets Markegång uppgår till lika belopp eller vswerstiger det salunda uträknade medium as scdnaste 5 Arens markegangspris, far ingen uppsägning till lefiverering tn Nalilia ega ruin,

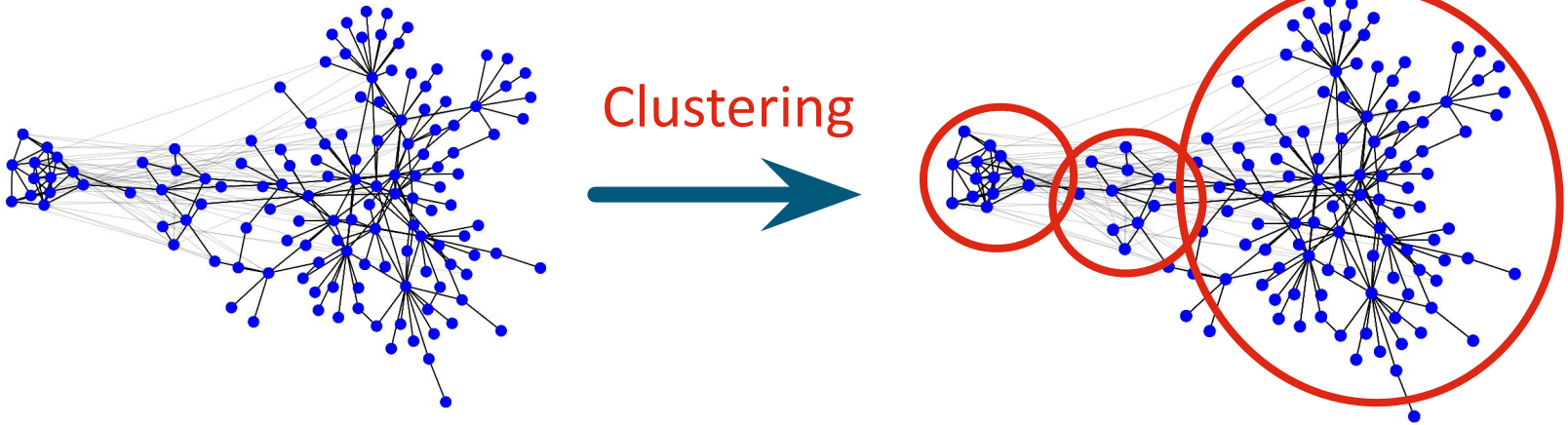
Graph view of the annotated data



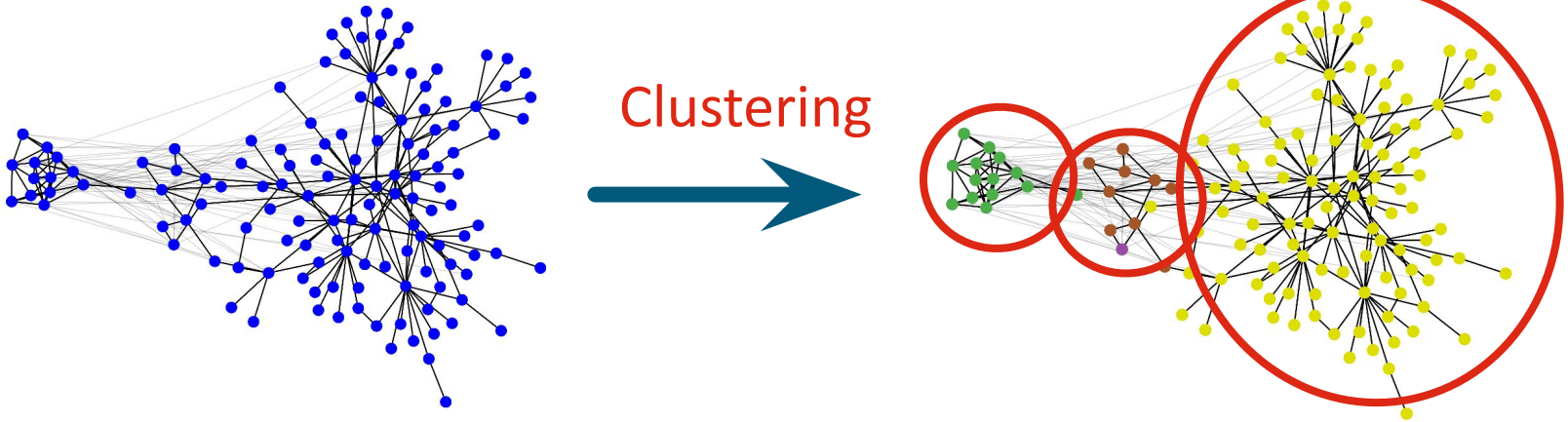
4: Identical
3: Closely related
2: Distantly related
1: Unrelated
Nodes from time T1
Nodes from time T2
No Edge = No annotation



How to measure change?



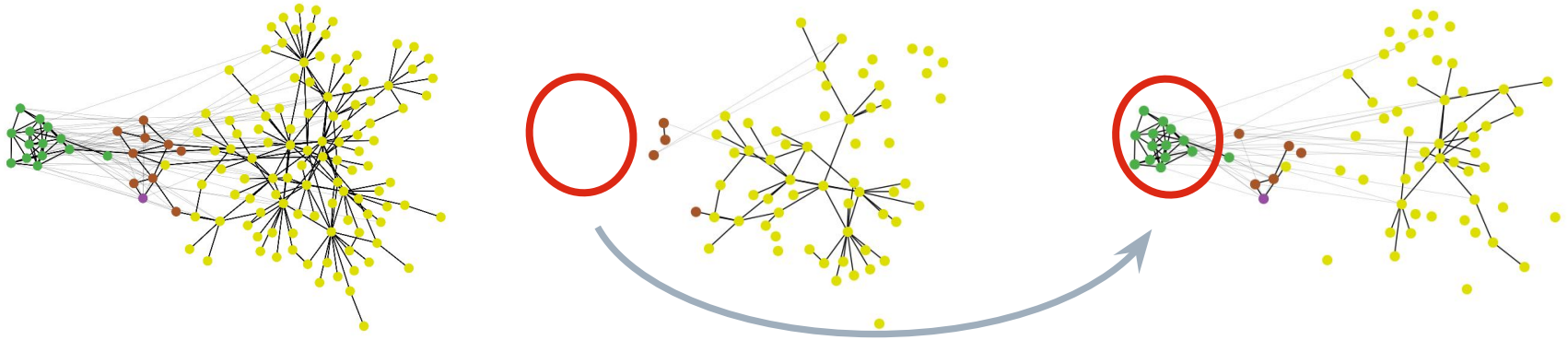
How to measure change?



Full clustered graph

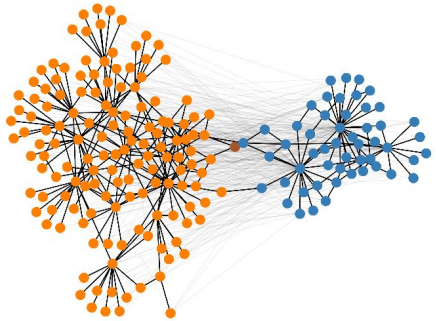
T1

T2



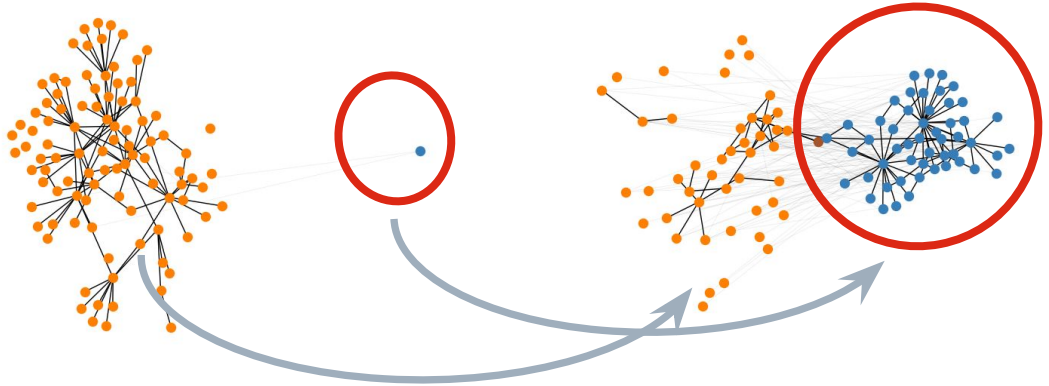
New cluster -> Binary change

Full clustered graph



T1

T2

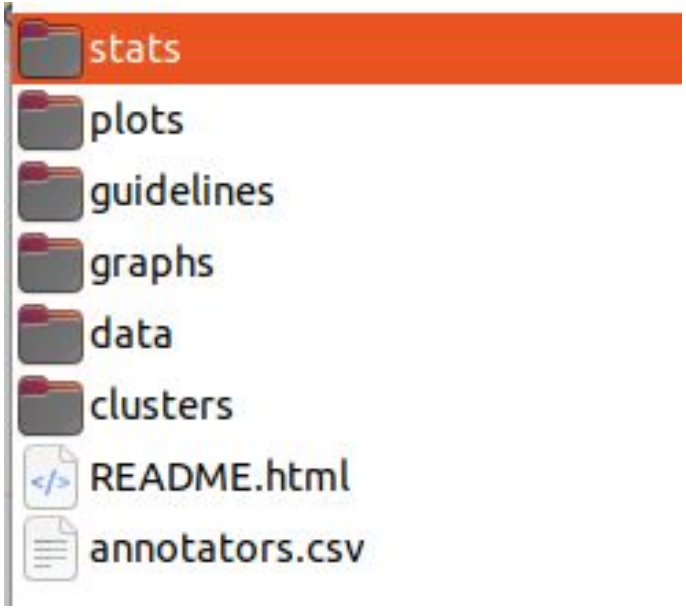


No new cluster
change in frequency of clusters
-> Graded change

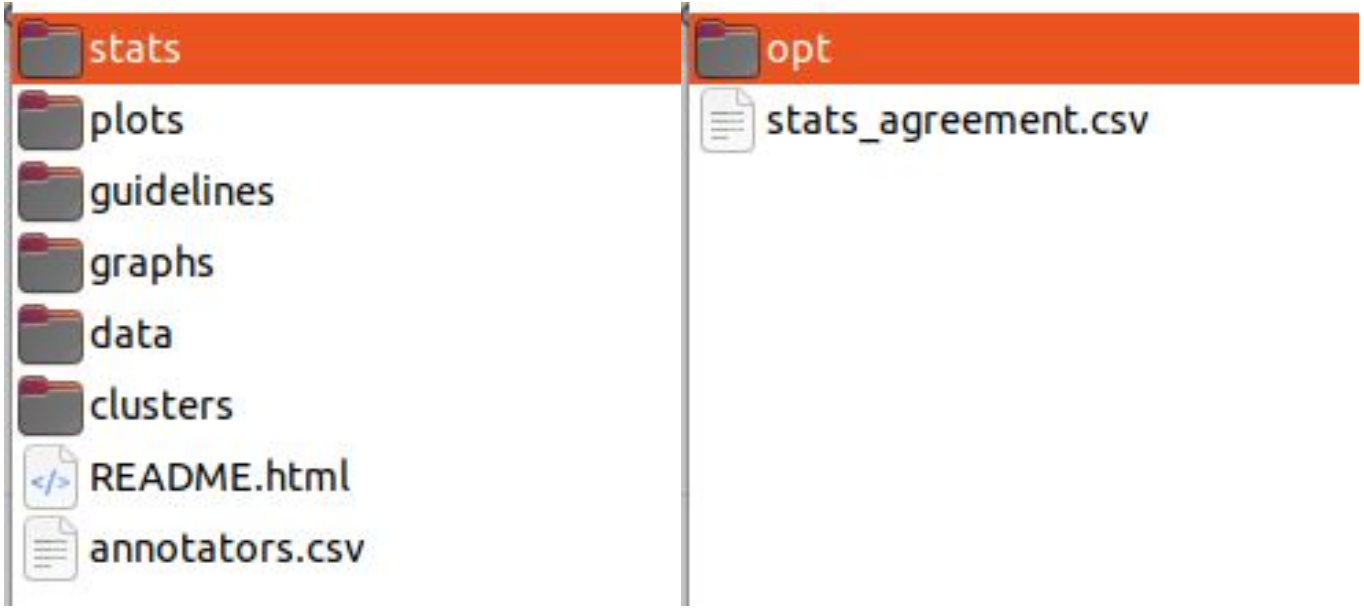
DWUGs

Language	Time periods	Diachronic Corpus	# targets
EN	C_1: 1810 – 1860 C_2: 1960 – 2010	C_1: CCOHA, C_2: CCOHA	46
SV	C_1: 1790 – 1830 C_2: 1895 – 1903	C_1: Kubhist, C_2: Kubhist	44
DE	C_1: 1800 – 1899 C_2: 1946 – 1990	C_1: DTA, C_2: BZ+ND	50
LA	C_1: 200 – 0 C_2: 0 – 2000	C_1: LatinISE, C_2: LatinISE	40
ES	C_1: 1810 – 1906 C_2: 1994 – 2020	C_1: PG, C_2: TED2013, NC, MultiUN, Europarl	100
RU	C_1: 1700 – 1916 C_2: 1918 – 1990 C_3: 1992 –2016	C_1, C_2, C_3: RNC	111
NO	C_1: 1929 –1965 C_2: 1970 – 2013	C_1: NBdigital, C_2: NBdigital	40
NO	C_1: 1980 – 1990 C_2: 2012 – 2019	C_1: NBdigital, C_2: NAK	40
ZH	C_1: 1954 – 1978 C_2: 1979 – 2003	C_1, C_2: People's Daily	40

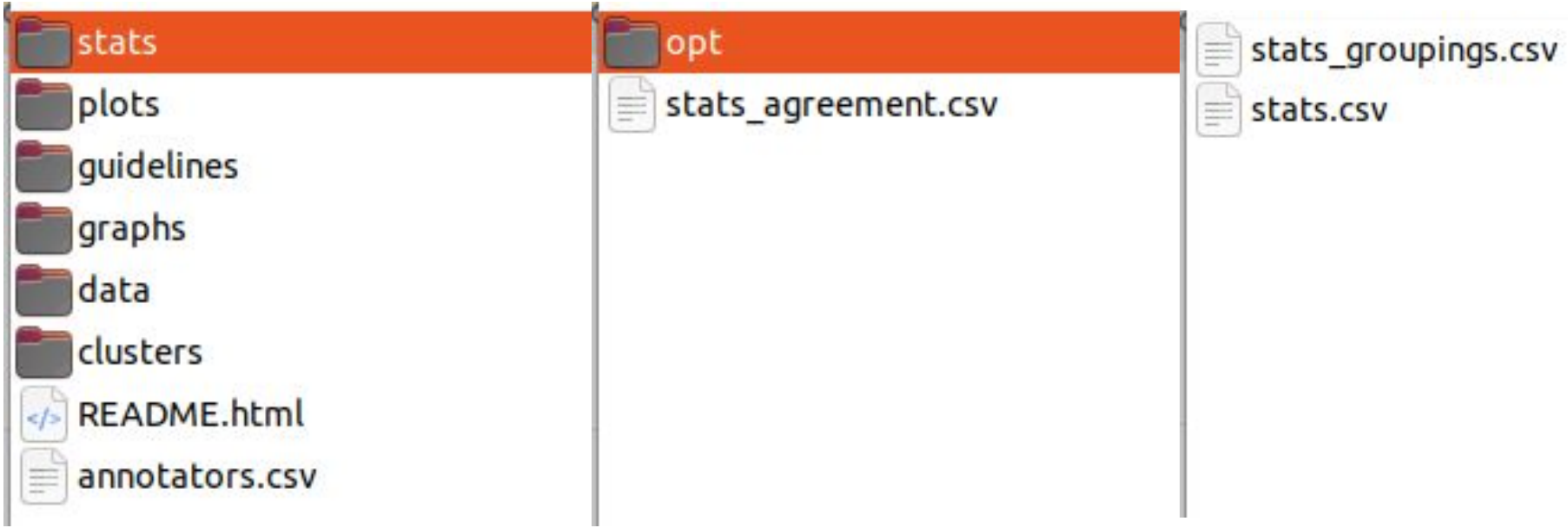
DURel Output



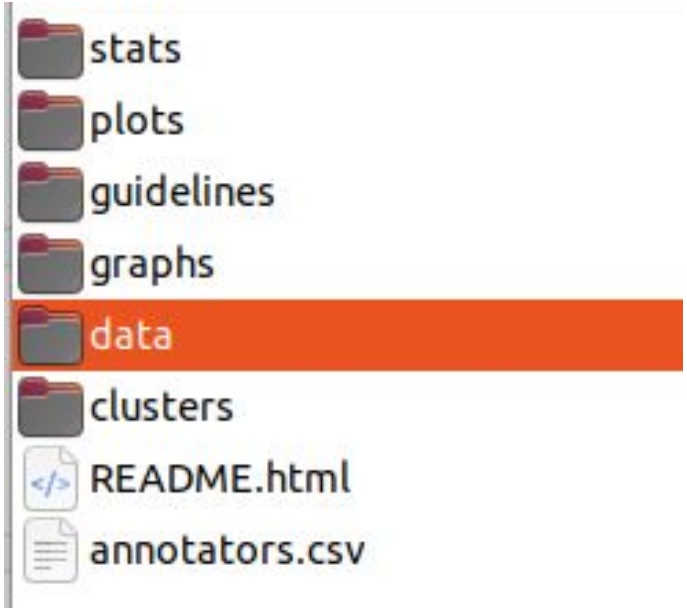
DURel Output



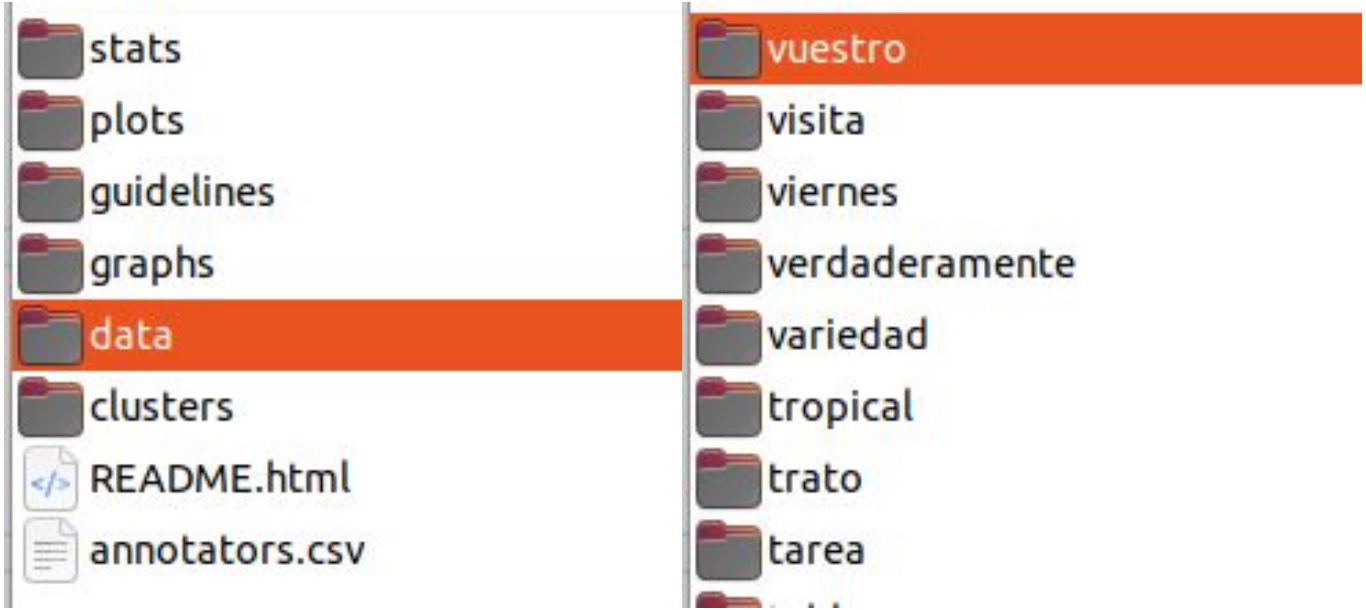
DURel Output



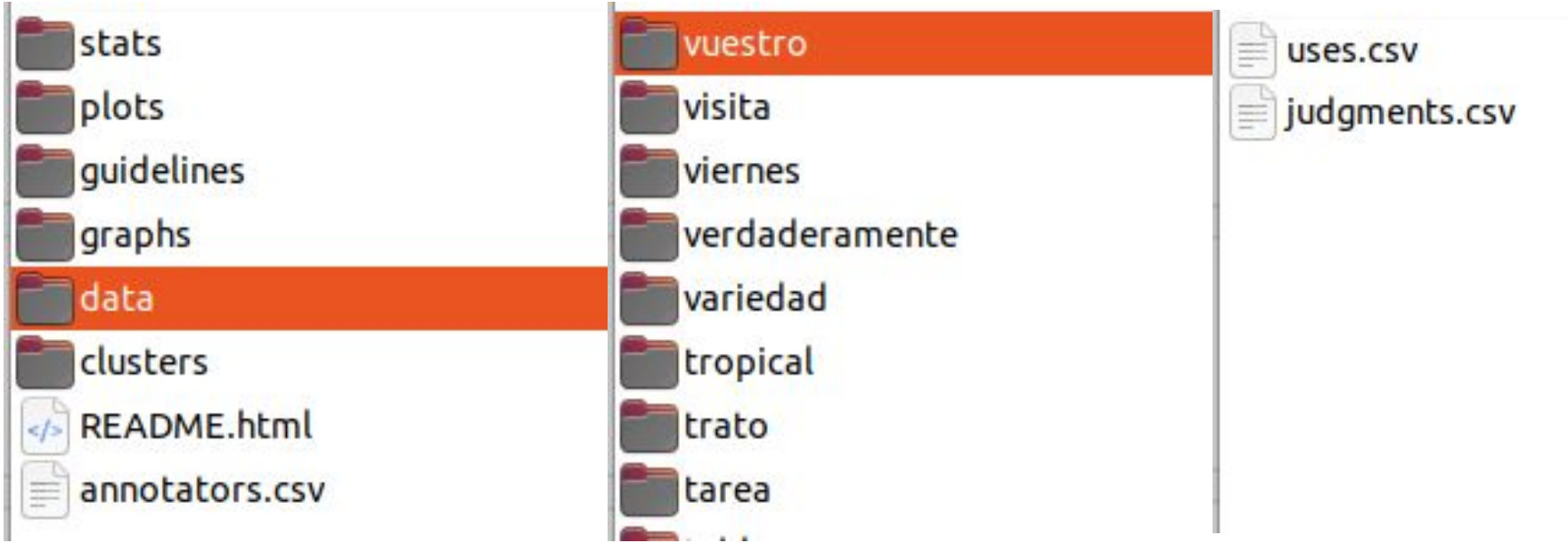
DURel Output



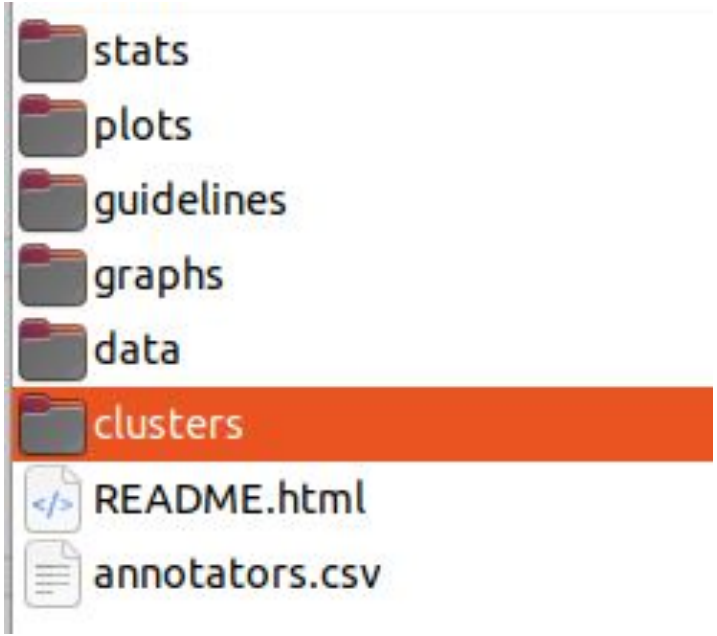
DURel Output



DURel Output



DURel Output



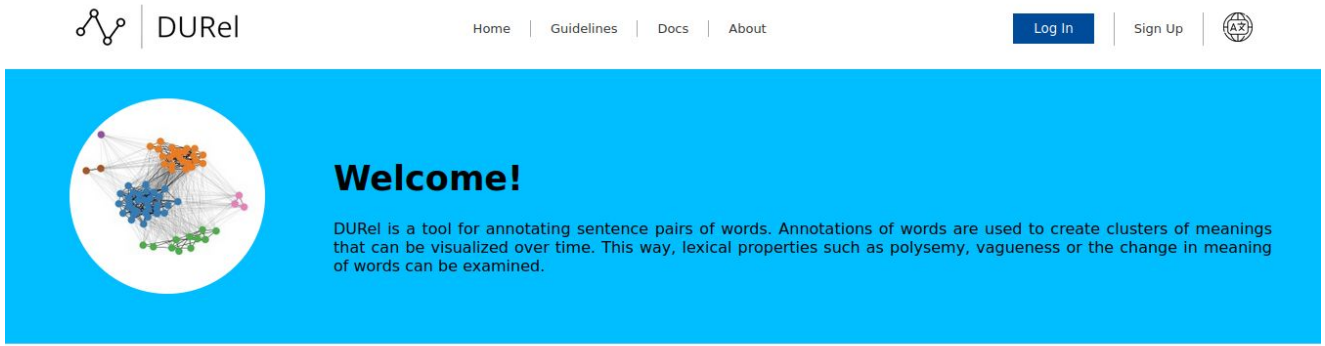
DURel Output

DURel Output

DURel Output

identifier	cluster
old_corpus_spanish-576808-788	0
old_corpus_spanish-532171-191	0
old_corpus_spanish-487991-132	0
old_corpus_spanish-530109-166	0
old_corpus_spanish-593902-84	0
old_corpus_spanish-408016-67	0
old_corpus_spanish-23203-53	1
old_corpus_spanish-146362-370	0

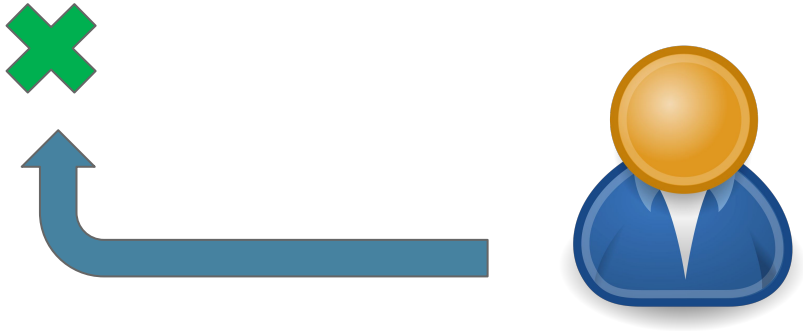
TRY DUREL!



https://durel.ims.uni-stuttgart.de/

DIACR-ITA

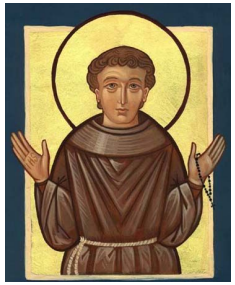
Roma, appalti **pilotati** e corruzione tra imprenditori e funzionari pubblici: 10 arresti



P Basile, A Caputo, T Caselli, P Cassotti, R Varvara. DIACR-Ita@ EVALITA2020: Overview of the evalita2020 diachronic lexical semantics (diacr-ita) task

SemEval 2020 Latin

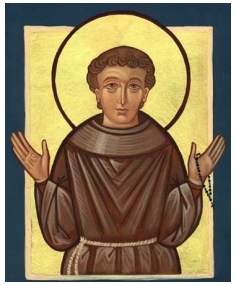
Cumque aliquandiu agitata esset praesens quaestio, dux, sicuti vir humilis erat, et mansuetus, ac timens sermones Domini, in die purificationis **beatae** Mariae, praesente clero et populo universo, ecclesiae Sanctae Resurrectionis, quartam partem Joppe resignavit.



4: Identical

SemEval 2020 Latin

Cumque aliquandiu agitata esset praesens quaestio, dux, sicuti vir humilis erat, et mansuetus, ac timens sermones Domini, in die purificationis **beatae** Mariae, praesente clero et populo universo, ecclesiae Sanctae Resurrectionis, quartam partem Joppe resignavit.



1: Unrelated

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling
- Human annotation of Lexical Semantic Change
- Diachronic Models of Language
  - **Static Models and Alignment**
  - Contextualized Models
  - Generative Models
- Hands-on

NAACL 2024

A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change

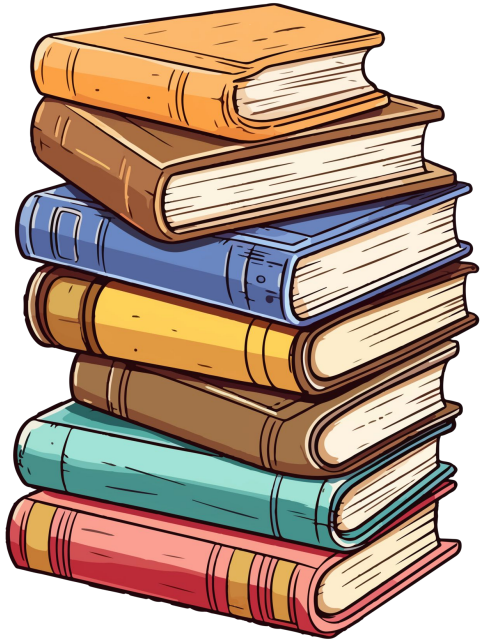
Change is Key!

Francesco Periti
francesco.periti@unimi.it

Nina Tahmasebi
nina.tahmasebi@gu.se

Lexical Semantic Change
Computational modeling

$$\mathcal{C} = \bigcup_{i=1}^{n} C_i$$



C_1 — 1700 – 1800

C_2 — 1800 – 1900

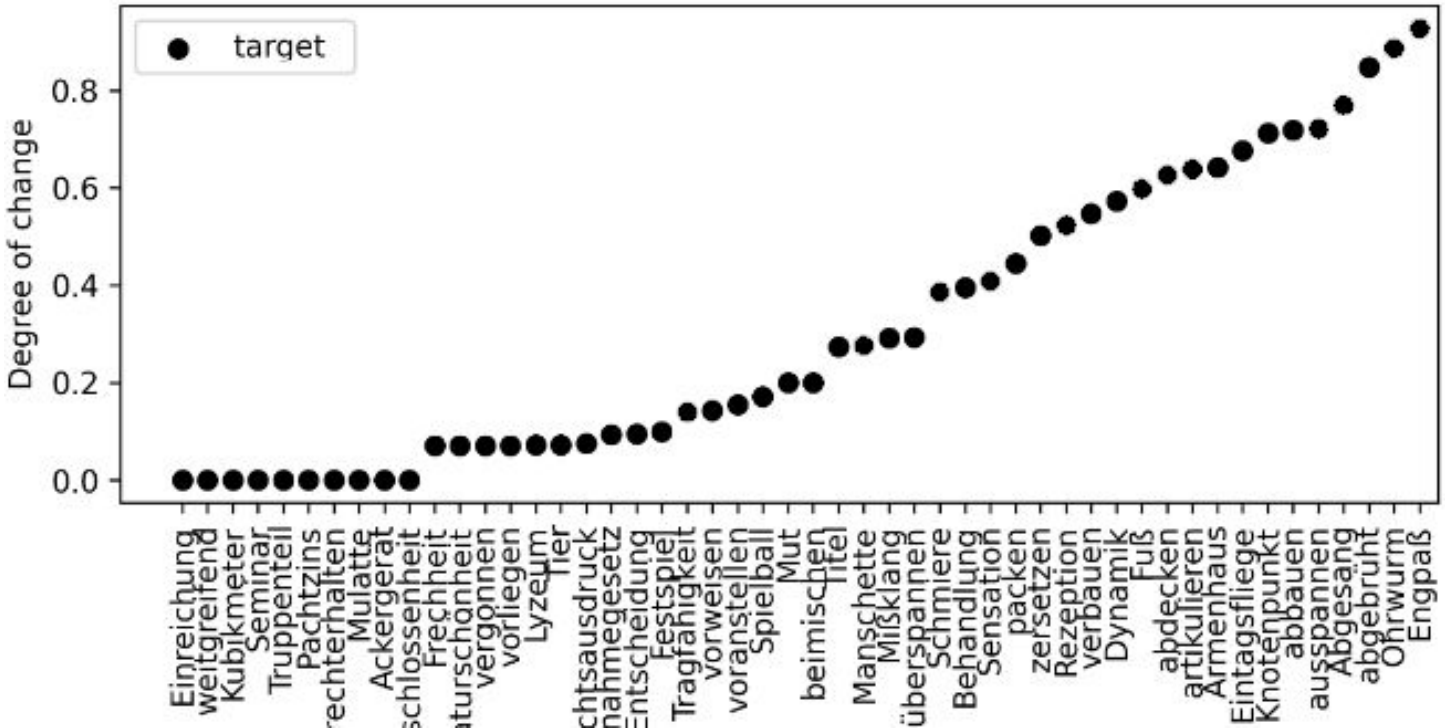
C_3 — 1900 – 2024

lexical semantic change

w `manufacture`
From *to make by hand*
To *to make by machine*

w `gay`
From *cheerful*
To *homosexual*

Lexical Semantic Change
Graded Change Detection



Issues

Memory consumption

Computation time

Different benchmarks

1. Unfair comparisons
2. Partial complexity

1st Experimental setup – Languages
Fair comparison on Graded Change Detection

44		Swedish (Schlechtweg et al., 2020)	$\langle t_1, t_2 \rangle$	40	 Norwegian (Kutuzov et al., 2022)
46		English (Schlechtweg et al., 2020)	$\langle t_1, t_2 \rangle$	40	 Chinese (Chen et al., 2023)
50		German (Schlechtweg et al., 2020)	$\langle t_1, t_2 \rangle$	100	 Russian (Kutuzov and Pivovarova., 2021)
40		Latin (Schlechtweg et al., 2020)	$\langle t_1, t_2 \rangle$	111	 Spanish (Zamora-Reina et al., 2022)

1st Experimental setup– Models
Fair comparison on Graded Change Detection

- Monolingual:
  - BERT (Devlin et al., 2019)
- Multilingual:
  - mBERT (Devlin et al., 2019)
  - XLM-R (Conneau et al., 2020)

1st Experimental setup – Layers
Fair comparison on Graded Change Detection

- Layers (Periti and Dubossarsky, 2023; Ma et al., 2019; Liang and Shi, 2023):
  - 1 … 12

1st Experimental setup – Approaches
Fair comparison on Graded Change Detection

- form-based:
  - `APD` (Giulianelli et al., 2020)
  - `PRT` (Martinc et al., 2020a)

- sense-based:
  - `AP+JSD` (Martinc et al., 2020b)
  - `WiDiD` (Periti et al., 2022)

1st Experimental setup – Approaches
Fair comparison on Graded Change Detection

- unsupervised:
  - `pre-trained LM + APD`
  - `pre-trained LM + PRT`
  - `pre-trained LM + AP+JSD`
  - `pre-trained LM + WiDiD`

- supervised:
  - `XL-LEXEME` (Cassotti et al., 2023) `+ APD`
  - `XL-LEXEME + PRT`
  - `XL-LEXEME + AP+JSD`
  - `XL-LEXEME + WiDiD`

Experimental results – Languages

Fair comparison on Graded Change Detection

	LMs	EN	LA	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
APD	BERT	.563	-	.271	.270	.335	.518	.482	.416	.441	.466	.656	.449
	mBERT	.363	.102	.398	.389	.341	.368	.345	.386	.279	.488	.689	.371
	XLM-R	.444	.151	.264	.257	.386	.290	.287	.318	.195	.379	.500	.316
	XL-LEXEME	**.886***	**.231**	**.839***	**.812***	**.665***	**.796***	**.820***	**.863***	**.659**	**.640***	**.731***	**.751***
PRT	BERT	.457	-	.422	.158	.413	.400	.374	.347	.507	.444	**.712**	.406
	mBERT	.270	.380	.436	.193	.543	.391	.356	.423	.219	.438	.524	.395
	XLM-R	.411	.424	.369	.020	.505	.321	.443	.405	.387	.149	.558	.381
	XL-LEXEME	**.676**	**.506***	**.824**	**.696**	**.632**	**.704**	**.750**	**.727**	**.764***	**.519**	.699	**.693**
AP+JSD	BERT	.289	-	.469	- .090	.225	.069	.279	.094	**.314**	.011	.165	.179
	mBERT	.181	.277	.280	.023	.067	.017	.086	- .116	.035	- .090	.465	.077
	XLM-R	.278	**.398**	.224	-.076	.224	- .068	**.209**	**.130**	- .100	.030	.448	.142
	XL-LEXEME	**.493**	.033	**.499**	**.118**	**.392**	**.106**	.053	.117	.297	**.381**	**.308**	**.223**
WiDiD	BERT	.385	-	.355	.106	.383	.135	.102	.243	.233	.087	.533	.239
	mBERT	.323	- .039	.312	.195	.343	- .068	.160	.142	.241	.290	.338	.181
	XLM-R	.564	- .064	.499	.129	.459	**.268**	.216	.342	.226	.349	.382	.314
	XL-LEXEME	**.652**	**.236**	**.677**	**.475**	**.522**	.178	**.354**	**.364**	**.561**	**.457**	**.563**	**.422**

Spearman's rank correlation coefficient

Experimental results – Form-based vs. Sense-based

Fair comparison on Graded Change Detection

	LMs	EN	LA	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
APD	BERT	.563	-	.271	.270	.335	.518	.482	.416	.441	.466	.656	.449
	mBERT	.363	.102	.398	.389	.341	.368	.345	.386	.279	.488	.689	.371
	XLM-R	.444	.151	.264	.257	.386	.290	.287	.318	.195	.379	.500	.316
	XL-LEXEME	**.886***	**.231**	**.839***	**.812***	**.665***	**.796***	**.820***	**.863***	**.659**	**.640***	**.731***	**.751***
PRT	BERT	.457	-	.422	.158	.413	.400	.374	.347	.507	.444	**.712**	.406
	mBERT	.270	.380	.436	.193	.543	.391	.356	.423	.219	.438	.524	.395
	XLM-R	.411	.424	.369	.020	.505	.321	.443	.405	.387	.149	.558	.381
	XL-LEXEME	**.676**	**.506***	**.824**	**.696**	**.632**	**.704**	**.750**	**.727**	**.764***	**.519**	.699	**.693**
AP+JSD	BERT	.289	-	.469	- .090	.225	.069	.279	.094	**.314**	.011	.165	.179
	mBERT	.181	.277	.280	.023	.067	.017	.086	- .116	.035	- .090	.465	.077
	XLM-R	.278	**.398**	.224	-.076	.224	- .068	**.209**	**.130**	- .100	.030	.448	.142
	XL-LEXEME	**.493**	.033	**.499**	**.118**	**.392**	**.106**	.053	.117	.297	**.381**	**.308**	**.223**
WiDiD	BERT	.385	-	.355	.106	.383	.135	.102	.243	.233	.087	.533	.239
	mBERT	.323	- .039	.312	.195	.343	- .068	.160	.142	.241	.290	.338	.181
	XLM-R	.564	- .064	.499	.129	.459	**.268**	.216	.342	.226	.349	.382	.314
	XL-LEXEME	**.652**	**.236**	**.677**	**.475**	**.522**	.178	**.354**	**.364**	**.561**	**.457**	**.563**	**.422**

Spearman's rank correlation coefficient

Experimental results – Monolingual vs. Multilingual

Fair comparison on Graded Change Detection

	LMs	EN	LA	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
APD	BERT	.563	-	.271	.270	.335	.518	.482	.416	.441	.466	.656	.449
	mBERT	.363	.102	.398	.389	.341	.368	.345	.386	.279	.488	.689	.371
	XLM-R	.444	.151	.264	.257	.386	.290	.287	.318	.195	.379	.500	.316
	XL-LEXEME	**.886***	**.231**	**.839***	**.812***	**.665***	**.796***	**.820***	**.863***	**.659**	**.640***	**.731***	**.751***
PRT	BERT	.457	-	.422	.158	.413	.400	.374	.347	.507	.444	**.712**	.406
	mBERT	.270	.380	.436	.193	.543	.391	.356	.423	.219	.438	.524	.395
	XLM-R	.411	.424	.369	.020	.505	.321	.443	.405	.387	.149	.558	.381
	XL-LEXEME	**.676**	**.506***	**.824**	**.696**	**.632**	**.704**	**.750**	**.727**	**.764***	**.519**	.699	**.693**
AP+JSD	BERT	.289	-	.469	- .090	.225	.069	.279	.094	**.314**	.011	.165	.179
	mBERT	.181	.277	.280	.023	.067	.017	.086	- .116	.035	- .090	.465	.077
	XLM-R	.278	**.398**	.224	-.076	.224	- .068	**.209**	**.130**	- .100	.030	.448	.142
	XL-LEXEME	**.493**	.033	**.499**	**.118**	**.392**	**.106**	.053	.117	.297	**.381**	**.308**	**.223**
WiDiD	BERT	.385	-	.355	.106	.383	.135	.102	.243	.233	.087	.533	.239
	mBERT	.323	- .039	.312	.195	.343	- .068	.160	.142	.241	.290	.338	.181
	XLM-R	.564	- .064	.499	.129	.459	**.268**	.216	.342	.226	.349	.382	.314
	XL-LEXEME	**.652**	**.236**	**.677**	**.475**	**.522**	.178	**.354**	**.364**	**.561**	**.457**	**.563**	**.422**

Spearman's rank correlation coefficient

Experimental results – Supervised vs. Unsupervised

Fair comparison on Graded Change Detection

	LMs	EN	LA	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
APD	BERT	.563	-	.271	.270	.335	.518	.482	.416	.441	.466	.656	.449
	mBERT	.363	.102	.398	.389	.341	.368	.345	.386	.279	.488	.689	.371
	XLM-R	.444	.151	.264	.257	.386	.290	.287	.318	.195	.379	.500	.316
	XL-LEXEME	**.886***	**.231**	**.839***	**.812***	**.665***	**.796***	**.820***	**.863***	**.659**	**.640***	**.731***	**.751***
PRT	BERT	.457	-	.422	.158	.413	.400	.374	.347	.507	.444	**.712**	.406
	mBERT	.270	.380	.436	.193	.543	.391	.356	.423	.219	.438	.524	.395
	XLM-R	.411	.424	.369	.020	.505	.321	.443	.405	.387	.149	.558	.381
	XL-LEXEME	**.676**	**.506***	**.824**	**.696**	**.632**	**.704**	**.750**	**.727**	**.764***	**.519**	.699	**.693**
AP+JSD	BERT	.289	-	.469	- .090	.225	.069	.279	.094	**.314**	.011	.165	.179
	mBERT	.181	.277	.280	.023	.067	.017	.086	- .116	.035	- .090	.465	.077
	XLM-R	.278	**.398**	.224	-.076	.224	- .068	**.209**	**.130**	- .100	.030	.448	.142
	XL-LEXEME	**.493**	.033	**.499**	**.118**	**.392**	**.106**	.053	.117	.297	**.381**	**.308**	**.223**
WiDiD	BERT	.385	-	.355	.106	.383	.135	.102	.243	.233	.087	.533	.239
	mBERT	.323	- .039	.312	.195	.343	- .068	.160	.142	.241	.290	.338	.181
	XLM-R	.564	- .064	.499	.129	.459	**.268**	.216	.342	.226	.349	.382	.314
	XL-LEXEME	**.652**	**.236**	**.677**	**.475**	**.522**	.178	**.354**	**.364**	**.561**	**.457**	**.563**	**.422**

Spearman's rank correlation coefficient

Experimental results – Layers

Fair comparison on Graded Change Detection

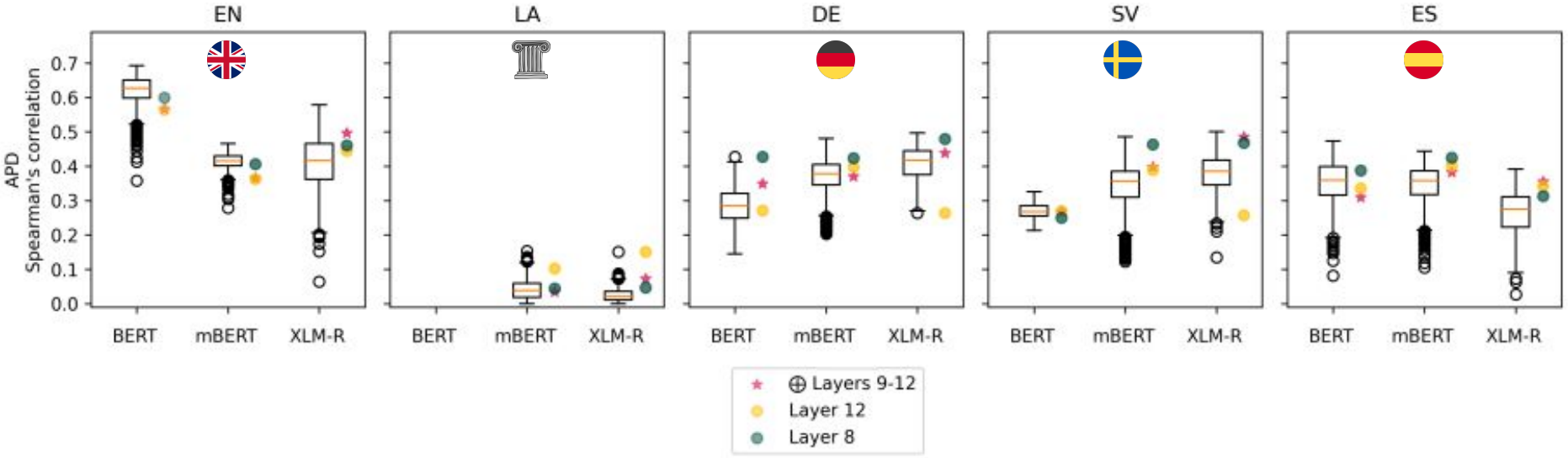


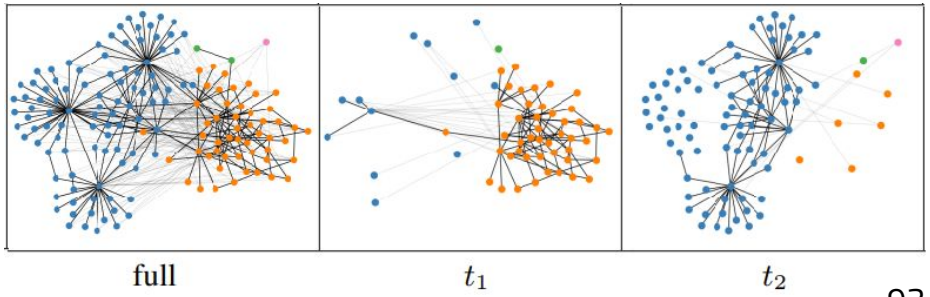
Figure 2: **Score distribution for GCD** obtained by using all possible layer combinations of length 2 (e.g., Layer 1 and 2), length 3 (e.g., Layer 10, 11, 12), and length 4 (e.g., Layer 1, 10, 11, 12) for BERT, mBERT, and XLM-R. The y-axis represents the Spearman correlation. We highlight the performance for GCD obtained using Layer 8, Layer 12, and the sum of the last 4 layers (i.e., ⊕ 9-12).

2nd Experimental setup
Full complexity of Lexical Semantic Change

Benchmark creation

- *Semantic proximity judgments* of word in-context
- *Word sense induction* based on proximity judgments
- *Quantification of semantic change* from induced senses

- c_1: But we are most familiar with the exhibitions of gravity in bodies descending inclined **planes**, as in the avalanche and the cataract.
- c_2: Over the next several years, he said, the Coast Guard will get 60 more people, two new 270-foot vessels and al twin-engine **planes**.



2nd Experimental setup
Full complexity of Lexical Semantic Change

Benchmark creation

- *Semantic proximity judgments* of word in-context
- *Word sense induction* based on proximity judgments
- *Quantification of semantic change* from induced senses

Computational annotators

- *Word-in-Context*
- *Word Sense Induction*
- *Graded Change Detection*

`BERT, mBERT, XLM-R, XL-LEXEME`
`GPT-4` (OpenAI, 2023)

Experimental results – Evaluation tasks
Full complexity of Lexical Semantic Change

		EN	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
WiC	BERT	.503	.350	.221	.319	.314	.344	.350	.429	.406	.516	.358
	mBERT	.332	.344	.284	.289	.280	.273	.293	.283	.333	.413	.301
	XLM-R	.352	.289	.255	.288	.212	.250	.251	.317	.261	.392	.272
	XL-LEXEME	**.626**	**.628**	**.631**	**.547**	**.549**	**.558**	**.564**	**.484**	**.521**	**.630**	**.568**
	GPT-4.0	.606	-	-	-	-	-	-	-	-	-	-
	Agreement	*.633*	*.666*	*.672*	*.531*	*.531*	*.567*	*.564*	*.761*	*.667*	*.602*	*.593*
WSI	BERT	.136 / .700	.047 / .662	.023 / .596	.189 / .695	- / -	- / -	- / -	.251 / .771	.247 / .758	.279 / .759	.166 / .702
	mBERT	.067 / .644	.054 / .679	.024 / .648	.228 / .700	- / -	- / -	- / -	.241 / .759	.159 / .753	.172 / .713	.146 / .696
	XLM-R	.068 / .737	.024 / .725	.031 / .680	.164 / .755	- / -	- / -	- / -	.179 / .775	.183 / .715	.279 / .806	.133 / .743
	XL-LEXEME	.273 / .834	**.300 / .788**	**.249 / .766**	**.400 / .820**	- / -	- / -	- / -	**.337 / .806**	**.304 / .808**	**.448 / .836**	**.339 / .810**
	GPT-4.0	**.340 / .877**	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
GCD	BERT	.425	.116	.148	.284	.487	.452	.469	.571	.521	**.808**	.422
	mBERT	.120	.205	.234	.394	.372	.325	.408	.290	.454	.737	.357
	XLM-R	.219	.069	.143	.464	.284	.301	.375	.395	.345	.557	.324
	XL-LEXEME	.801	**.799**	**.721**	**.655**	**.780**	**.824**	**.851**	**.620**	**.567**	.716	**.754**
	GPT-4.0	**.818**	-	-	-	-	-	-	-	-	-	-

WiC: Spearman's correlation – **WSI:** ARI / PURITY – **GCD:** Spearman's correlation

Experimental results – Evaluation tasks
Full complexity of Lexical Semantic Change

		EN	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
WiC	BERT	.503	.350	.221	.319	.314	.344	.350	.429	.406	.516	.358
	mBERT	.332	.344	.284	.289	.280	.273	.293	.283	.333	.413	.301
	XLM-R	.352	.289	.255	.288	.212	.250	.251	.317	.261	.392	.272
	XL-LEXEME	.626	.628	.631	.547	.549	.558	.564	.484	.521	.630	.568
	GPT-4.0	.606	-	-	-	-	-	-	-	-	-	-
	Agreement	*.633*	*.666*	*.672*	*.531*	*.531*	*.567*	*.564*	*.761*	*.667*	*.602*	*.593*
WSI	BERT	.136 / .700	.047 / .662	.023 / .596	.189 / .695	- / -	- / -	- / -	.251 / .771	.247 / .758	.279 / .759	.166 / .702
	mBERT	.067 / .644	.054 / .679	.024 / .648	.228 / .700	- / -	- / -	- / -	.241 / .759	.159 / .753	.172 / .713	.146 / .696
	XLM-R	.068 / .737	.024 / .725	.031 / .680	.164 / .755	- / -	- / -	- / -	.179 / .775	.183 / .715	.279 / .806	.133 / .743
	XL-LEXEME	.273 / .834	**.300 / .788**	**.249 / .766**	**.400 / .820**	- / -	- / -	- / -	**.337 / .806**	**.304 / .808**	**.448 / .836**	**.339 / .810**
	GPT-4.0	**.340 / .877**	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
GCD	BERT	.425	.116	.148	.284	.487	.452	.469	.571	.521	.808	.422
	mBERT	.120	.205	.234	.394	.372	.325	.408	.290	.454	.737	.357
	XLM-R	.219	.069	.143	.464	.284	.301	.375	.395	.345	.557	.324
	XL-LEXEME	.801	**.799**	**.721**	**.655**	**.780**	**.824**	**.851**	**.620**	**.567**	.716	**.754**
	GPT-4.0	**.818**	-	-	-	-	-	-	-	-	-	-

WiC: Spearman's correlation – **WSI:** ARI / PURITY – **GCD:** Spearman's correlation

Experimental results – Evaluation tasks
Full complexity of Lexical Semantic Change

		EN	DE	SV	ES	RU			NO		ZH	Avg_w
		$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
WiC	BERT	.503	.350	.221	.319	.314	.344	.350	.429	.406	.516	.358
	mBERT	.332	.344	.284	.289	.280	.273	.293	.283	.333	.413	.301
	XLM-R	.352	.289	.255	.288	.212	.250	.251	.317	.261	.392	.272
	XL-LEXEME	.626	.628	.631	.547	.549	.558	.564	.484	.521	.630	.568
	GPT-4.0	.606	-	-	-	-	-	-	-	-	-	-
	Agreement	*.633*	*.666*	*.672*	*.531*	*.531*	*.567*	*.564*	*.761*	*.667*	*.602*	*.593*
WSI	BERT	.136 / .700	.047 / .662	.023 / .596	.189 / .695	- / -	- / -	- / -	.251 / .771	.247 / .758	.279 / .759	.166 / .702
	mBERT	.067 / .644	.054 / .679	.024 / .648	.228 / .700	- / -	- / -	- / -	.241 / .759	.159 / .753	.172 / .713	.146 / .696
	XLM-R	.068 / .737	.024 / .725	.031 / .680	.164 / .755	- / -	- / -	- / -	.179 / .775	.183 / .715	.279 / .806	.133 / .743
	XL-LEXEME	.273 / .834	.300 / .788	.249 / .766	.400 / .820	- / -	- / -	- / -	.337 / .806	.304 / .808	.448 / .836	.339 / .810
	GPT-4.0	.340 / .877	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
GCD	BERT	.425	.116	.148	.284	.487	.452	.469	.571	.521	**.808**	.422
	mBERT	.120	.205	.234	.394	.372	.325	.408	.290	.454	.737	.357
	XLM-R	.219	.069	.143	.464	.284	.301	.375	.395	.345	.557	.324
	XL-LEXEME	.801	**.799**	**.721**	**.655**	**.780**	**.824**	**.851**	**.620**	**.567**	.716	**.754**
	GPT-4.0	**.818**	-	-	-	-	-	-	-	-	-	-

WiC: Spearman's correlation – **WSI:** ARI / PURITY – **GCD:** Spearman's correlation

Concluding remarks
Future directions

- A systematic evaluation of LMs and approaches for modeling LSC.
  - obtaining state-of-the-art performance on Graded Change Detection does not solve the Lexical Semantic Change problem
  - there is a need to distinguish the different senses of a word

- A systematic evaluation of LMs as computational annotators.
  - models can be used to aid the evaluation of word meaning in context
  - more consistent approaches are needed to cluster word senses

Thank for your attention!

Change Is Key!

https://www.changeiskey.org/

pierluigi.cassotti@gu.se