The background features a blurred image of a bookshelf filled with books. A semi-transparent blue overlay covers the entire image. A horizontal orange bar is positioned across the middle of the page, behind the main text.

# Hands-on Day 3: DWUGs and measures of change

Pierluigi Cassotti, [pierluigi.cassotti@gu.se](mailto: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](mailto:pierluigi.cassotti@gu.se)

# Hands-on

# Hands-on



Colab Notebook

# 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

# DWUGs - English

```
from languagechange.benchmark import DWUG
```

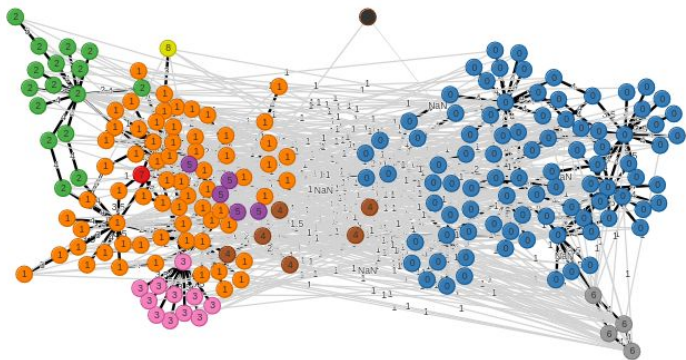
```
dwug_en = DWUG('EN', '2.0.1')
```

# DWUGs - English

```
from languagechange.benchmark import DWUG
```

```
dwug_en = DWUG('EN', '2.0.1')
```

```
dwug_en.show_usage_graph('plane_nn')
```

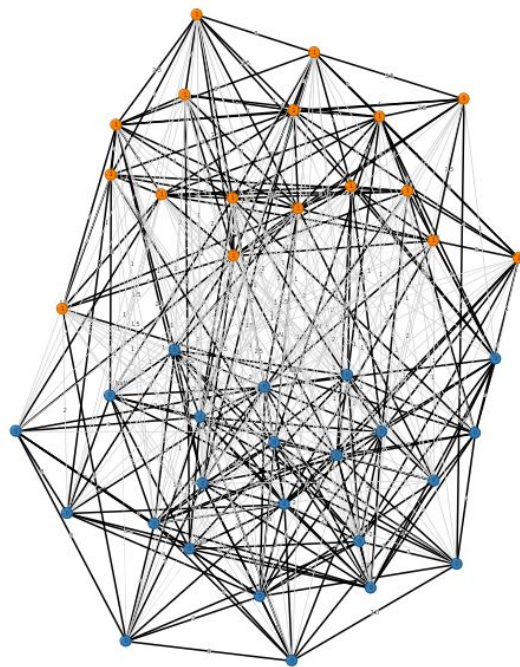


# DWUGs - Spanish

```
from languagechange.benchmark import DWUG
```

```
dwug_en = DWUG('ES', '4.0.0 (full)')
```

```
dwug_en.show_usage_graph('servidor')
```





# Computational representation of lexical items

```
from languagechange.models.representation.contextualized import XL_LEXEME
```

```
model = XL_LEXEME(device='cuda')
```

# Computational representation of lexical items

```
from languagechange.benchmark import DWUG
```

```
word = 'plane_nn'
```

```
usages_time1 = dwug_en.get_word_usages(word,group='1')
```

```
usages_time2 = dwug_en.get_word_usages(word,group='2')
```

```
vectors_time1 = model.encode(usages_time1)
```

```
vectors_time2 = model.encode(usages_time2)
```

# Computational representation of meaning

```
from languagechange.benchmark import DWUG
from languagechange.models.representation.contextualized import XL_LEXEME
from sklearn.cluster import AgglomerativeClustering
```

```
word = 'graft_nn'
```

```
usages_time1 = dwug_en.get_word_usages(word, group='1')
```

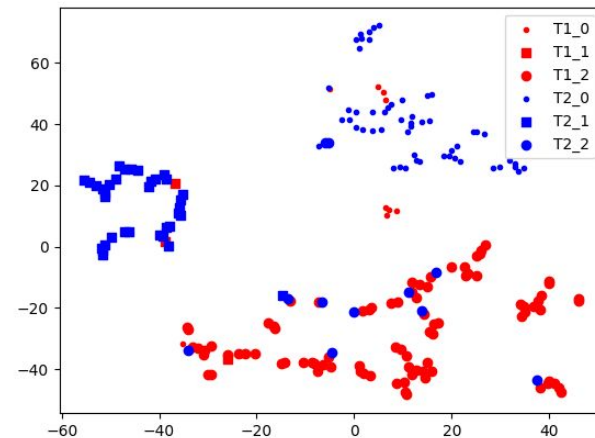
```
usages_time2 = dwug_en.get_word_usages(word, group='2')
```

```
vectors_time1 = model.encode(usages_time1)
```

```
vectors_time2 = model.encode(usages_time2)
```

```
vectors = np.concatenate((vectors_time1, vectors_time2), axis=0)
```

```
clustering=AgglomerativeClustering(n_clusters=None, linkage='average', metric='cosine', distance_threshold=0.5).fit(vectors)
```



# Computational req

```
from languagechange.benchmark import  
from languagechange.models.repre  
from sklearn.cluster import Aggl
```

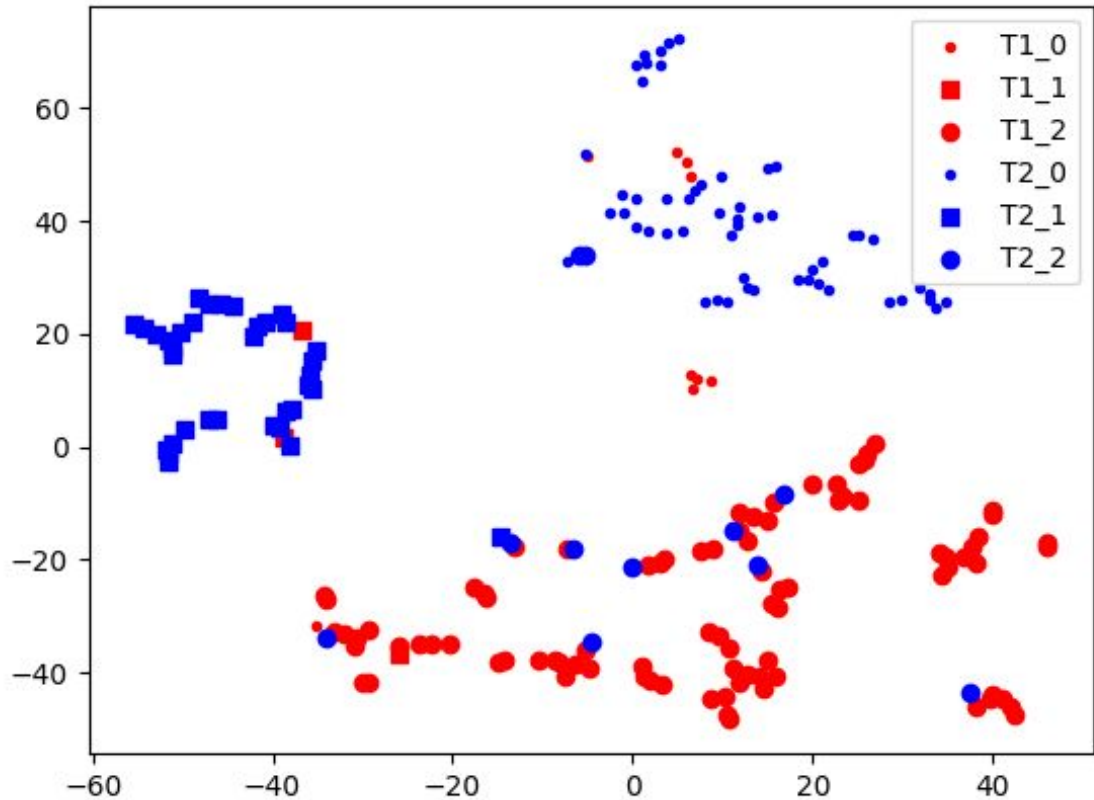
```
word = 'graft_nn'
```

```
usages_time1 = dwug_en.get_word_  
usages_time2 = dwug_en.get_word_
```

```
vectors_time1 = model.encode(usa  
vectors_time2 = model.encode(usa
```

```
vectors = np.concatenate((vector
```

```
clustering=AgglomerativeClusteri  
0.5).fit(vectors)
```



# Computational representation of meaning

CLUSTER 0

T1 " Mr. STOMS asks: ' Why **graft** on the quince stock at all? '

T1 The influence of the **graft** on the stock seems scarcely to extend beyond the power of communicating disease.

T1 Another perpendicular stroke slits the stock for the **graft**, leaving a perfectly smooth face cut for its reception.

T1 \* The influence exerted by the **graft** on the stock, although not strictly within the limits of thi chapter, offers an interesting subject for inquiry.

T1 The wedge b opens the stock to receive the **graft**.

# Computational representation of meaning

## CLUSTER 0

T1 " Mr. STOMS asks: ' Why **graft** on the quince stock at all? '

T1 The influence of the **graft** on the stock seems scarcely to extend beyond the power of communicating disease.

T1 Another perpendicular stroke slits the stock for the **graft**, leaving a perfectly smooth face cut for its reception.

T1 \* The influence exerted by the **graft** on the stock, although not strictly within the limits of this chapter, offers an interesting subject for inquiry.

## CLUSTER 1

T1 The we T1 When the **graft** is united and begins to make a growth, the bandage

T1 Indeed, it is a mooted question whether **grafts** should be taken at all from such a source; but I see no reason for going to this extreme.

T1 disposition, and is somewhat forgetful of qwq cealed a great deal of truth: " The setter is a product of art, as much as the Queen Claude Ilium or double rose; he is a dumb dog, grafted on the running dog, and which returns to the wild stock, like the double rose, when the **graft** fails to take effect. "

T2 He asked what evidence there was that the " immunity " so commonly spoken of was an immunity directed against the tumor as such, or whether it might not merely be an immunity directed against the tumor **graft** considered as a genetically foreign cell?

T2 The entire dosage regimen is a knife edge in which underdosage will lead to the rejection of the **graft** and overdosage to secondary damage on cells other than those that transact immunological responses?

T2 Lung transplants done in 32 patients during the last ten years have been quite disappointing, with only 3 patients living for more than thirty days | with a functioning **graft**.

# Computational representation of meaning

CLUSTER 0

T1 " Mr. STOMS asks: ' Why **graft** on the quince stock at all? '

T1 The influence of the **graft** on the stock seems scarcely to extend beyond the power of communicating disease.

T1 Another perpendicular stroke slits the stock for the **graft**, leaving a perfectly smooth face cut for its reception.

T1 \* The influence exerted by the **graft** on the stock, although not strictly within the limits of thi chapter, offers an interesting subject for inquiry.

T1 The wedge b opens the stock to receive the **graft**.

T1 When the **graft** is united and begins to make a growth, the bandage

T1 Indeed, it is a mooted question whether **grafts** should be taken at all from such a source; but I see no reason for going to this extreme.

T1 disposition, and is somewhat forgetful of qwq cealed a great deal of truth: " The setter is a product of art, as much as the Queen Claude Ilium or double rose; he is a dumb dog, grafted on the running dog, and which returns to the wild stock, like the double rose, when the **graft** fails to take effect. "

T2 He ask  
whether it  
CLUSTER 2

T1 A maiden plant (a tree but one year from the **graft**.) being planted " is to be headed down to four buds or eyes

T2 The en  
on cells ot  
T1 When the operation is performed close to the ground, or even under the surface; after the **graft** is properly tied draw up earth all round, leaving one or two buds above; this method is both convenient, and in many instances preferable, and more to be depended upon.

T2 Lung tr  
days | witi  
-- . . .  
T1 Fine mould or peat, from the box in which the roots were packed, is sprinkled over the bottom, the **grafts** are placed in, slanting, in successive layers, and all the spaces filled with mould or peat.

T1 For the same reason, root-grafting the pear is not adopted, especially as slightly unfavorable causes are apt to produce far greater failures such **grafts** than with the apple.

T1 In the following spring, the grafted trees may be looked over, and any young suckers growing from the stock or tree that may be liable to injure and retard the growth of the **graft**, may be taken off.

# LSC metrics: APD

```
from languagechange.models.change.metrics import APD, PRT, PJSD
```

```
word = 'plane_nn'
```

```
usages_time1 = dwug_en.get_word_usages(word, group='1')
```

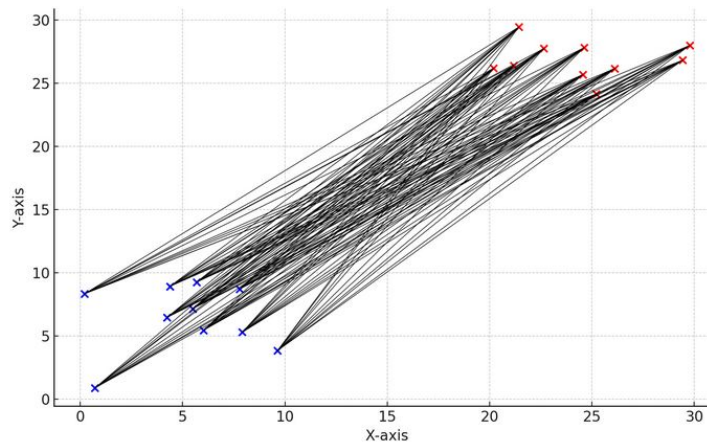
```
usages_time2 = dwug_en.get_word_usages(word, group='2')
```

```
vectors_time1 = model.encode(usages_time1)
```

```
vectors_time2 = model.encode(usages_time2)
```

```
apd_metric = APD()
```

```
apd_metric.compute_scores(vectors_time1, vectors_time2)
```





# LSC metrics: PRT

```
from languagechange.models.change.metrics import APD, PRT, PJSD
```

```
word = 'plane_nn'
```

```
usages_time1 = dwug_en.get_word_usages(word,group='1')
```

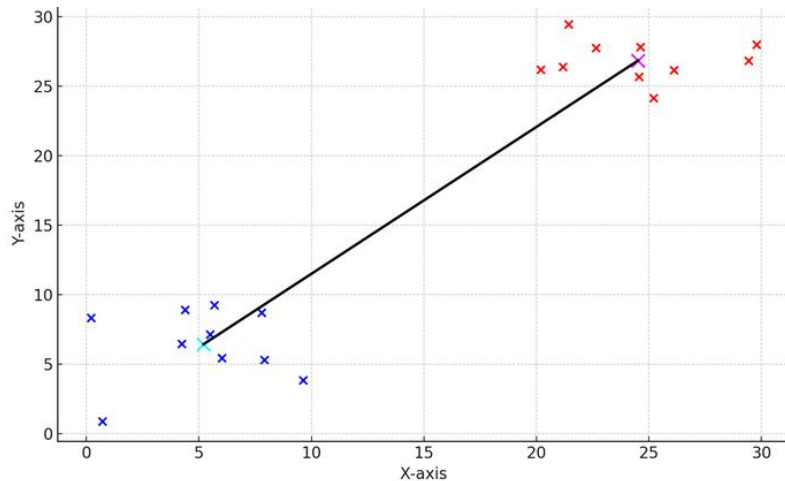
```
usages_time2 = dwug_en.get_word_usages(word,group='2')
```

```
vectors_time1 = model.encode(usages_time1)
```

```
vectors_time2 = model.encode(usages_time2)
```

```
prt_metric = PRT()
```

```
prt_metric.compute_scores(vectors_time1,vectors_time2)
```



# LSC metrics: Jensen-Shannon distance

```
from languagechange.models.change.metrics import APD, PRT, PJSD
```

```
word = 'plane_nn'
```

```
usages_time1 = dwug_en.get_word_usages(word, group='1')
```

```
usages_time2 = dwug_en.get_word_usages(word, group='2')
```

```
vectors_time1 = model.encode(usages_time1)
```

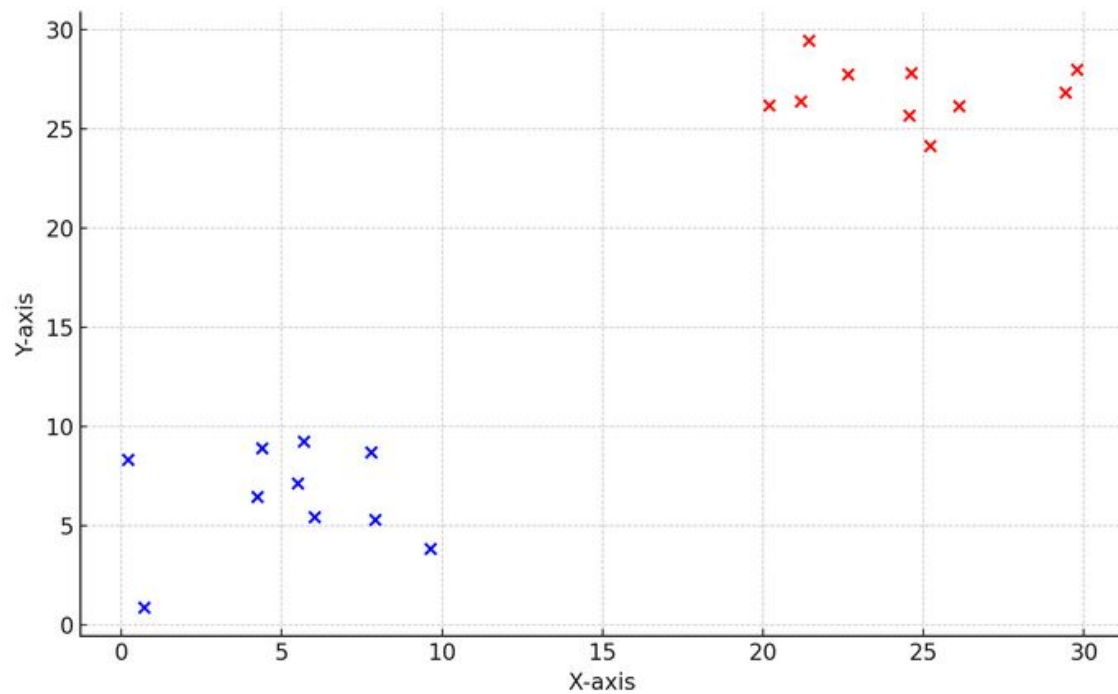
```
vectors_time2 = model.encode(usages_time2)
```

```
clustering=AgglomerativeClustering(n_clusters=None, linkage='average', metric='cosine', distance  
_threshold=0.5)
```

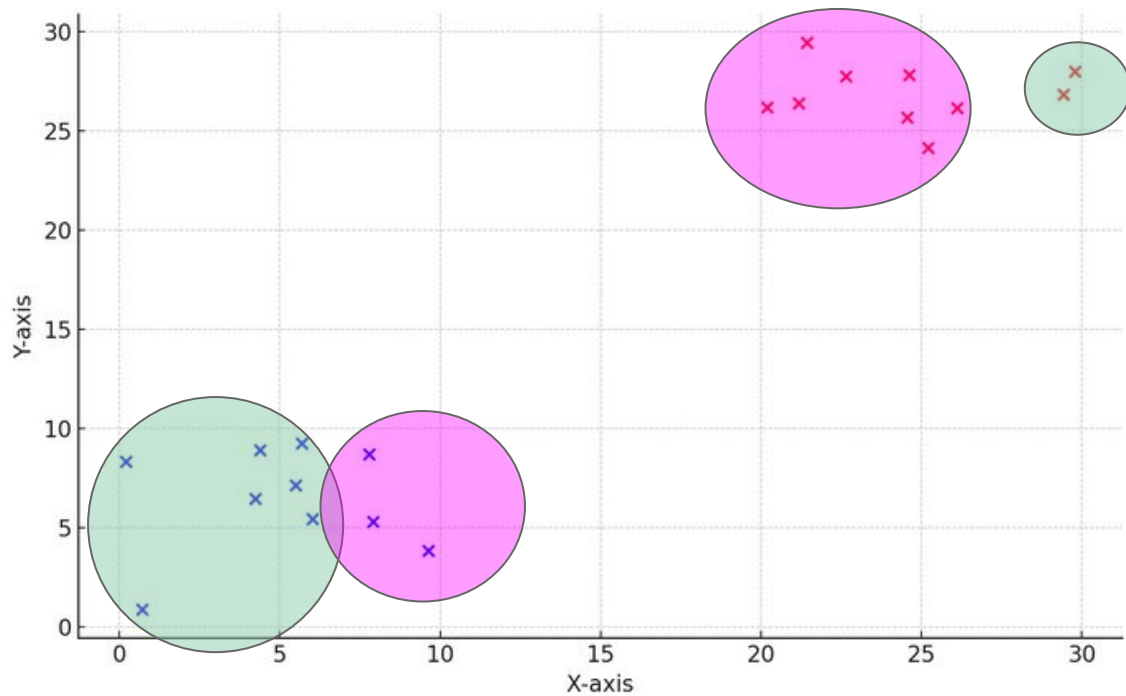
```
pjsd_metric = PJSD()
```

```
pjsd_metric.compute_scores(vectors_time1, vectors_time2, clustering)
```

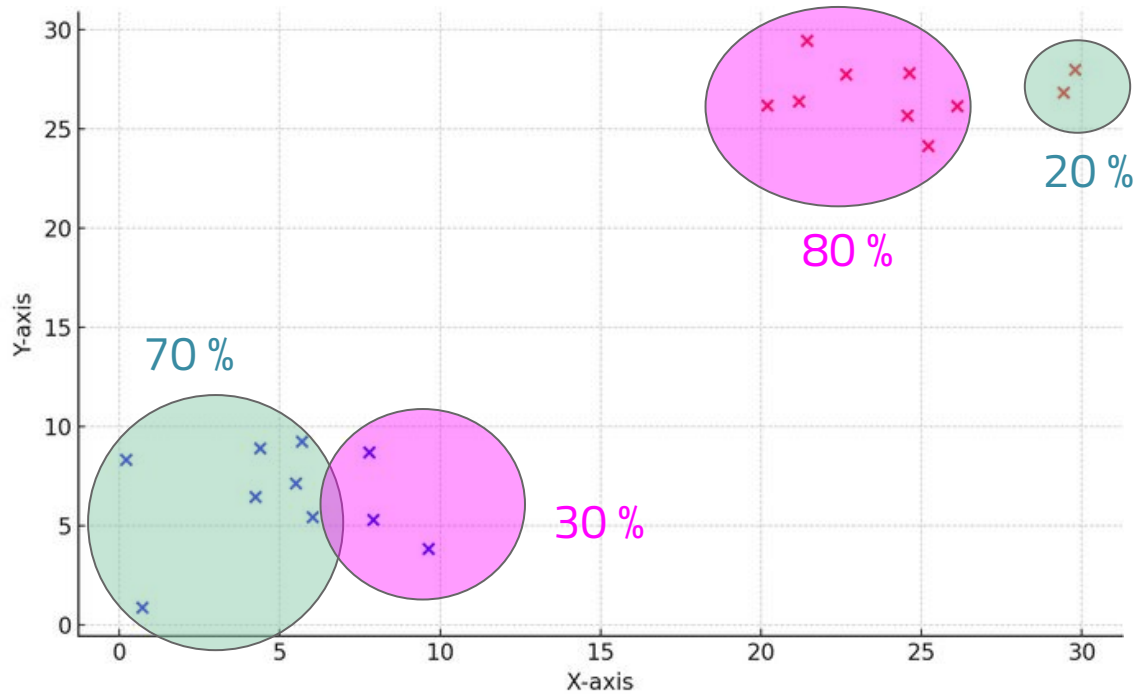
# LSC metrics: PJSD



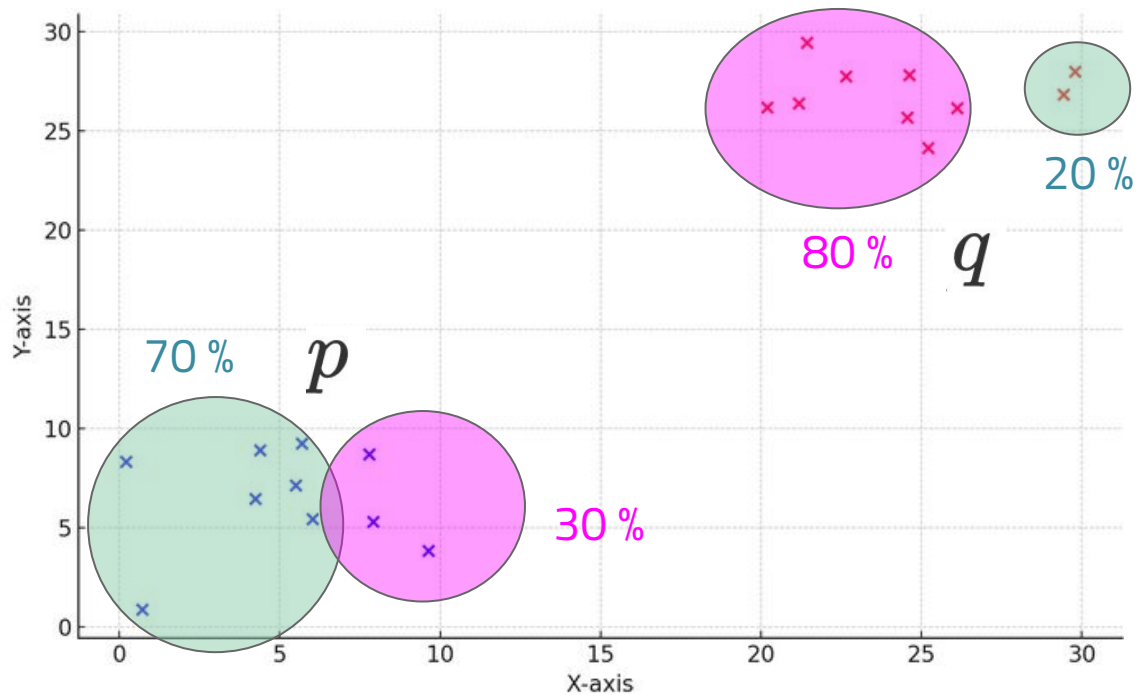
# LSC metrics: PJSD (Stage 1: Clustering)



# LSC metrics: PJSD (Stage 2: Clustering Distributions)



# LSC metrics: PJSD (Stage 2: Jensen-Shannon distance)



## LSC metrics: PJSD (Stage 2: Jensen-Shannon distance)

$$\sqrt{\frac{D(p \parallel m) + D(q \parallel m)}{2}}$$

$$m = \frac{1}{2}(P + Q)$$

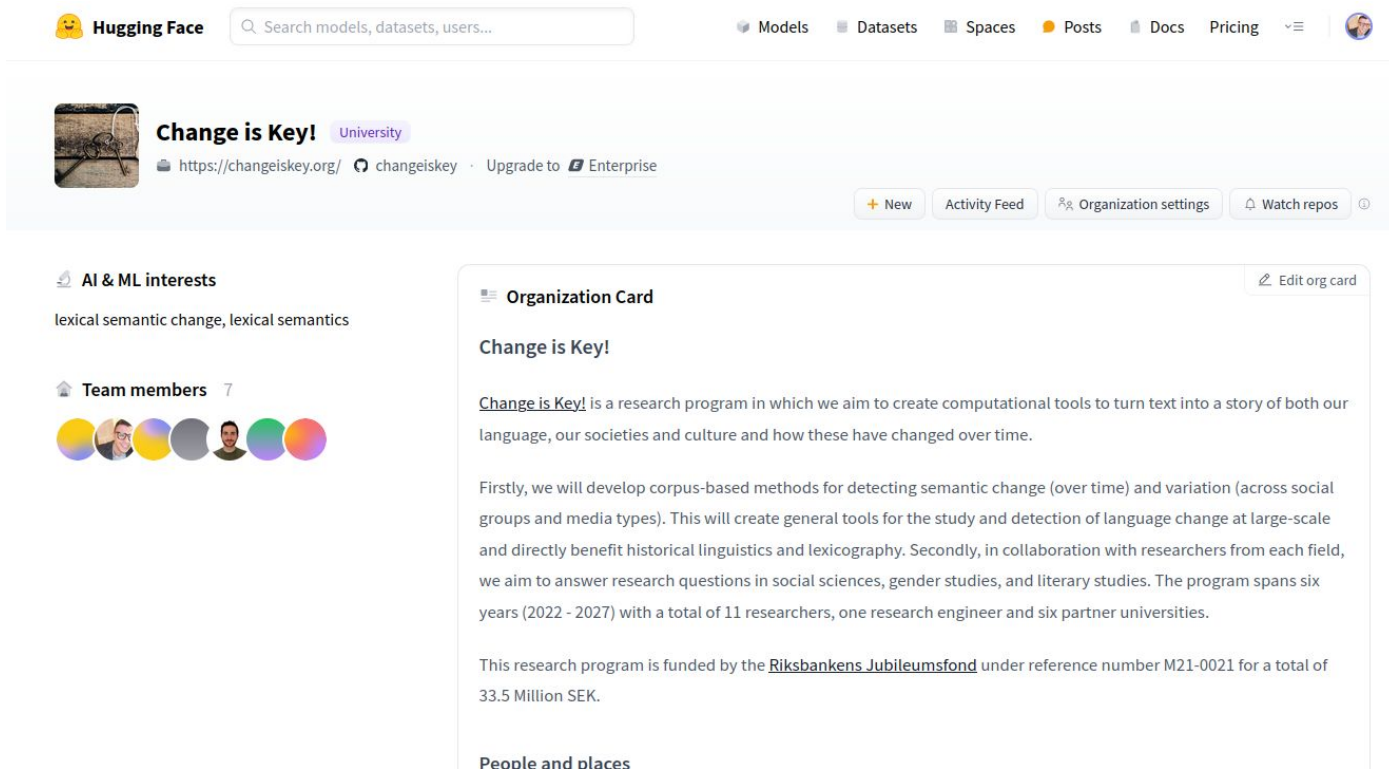
**GitHub**



[github.com/pierluigic/languagechange](https://github.com/pierluigic/languagechange)



# Change is Key! is on Huggingface



The screenshot shows the Hugging Face profile page for the organization 'Change is Key!'. The page includes a navigation bar with 'Models', 'Datasets', 'Spaces', 'Posts', 'Docs', and 'Pricing'. The organization's profile features a profile picture, the name 'Change is Key!' with a 'University' badge, and a link to their website. Below the profile, there are sections for 'AI & ML interests' (lexical semantic change, lexical semantics) and 'Team members' (7 members). The main content area displays an 'Organization Card' for 'Change is Key!' with a description of their research program and funding information.

**Hugging Face** Search models, datasets, users...

Models Datasets Spaces Posts Docs Pricing

**Change is Key!** University  
<https://changeiskey.org/> changeiskey · Upgrade to Enterprise

New Activity Feed Organization settings Watch repos

**AI & ML interests**  
lexical semantic change, lexical semantics

**Team members** 7

**Organization Card** Edit org card

**Change is Key!**

Change is Key! is a research program in which we aim to create computational tools to turn text into a story of both our language, our societies and culture and how these have changed over time.

Firstly, we will develop corpus-based methods for detecting semantic change (over time) and variation (across social groups and media types). This will create general tools for the study and detection of language change at large-scale and directly benefit historical linguistics and lexicography. Secondly, in collaboration with researchers from each field, we aim to answer research questions in social sciences, gender studies, and literary studies. The program spans six years (2022 - 2027) with a total of 11 researchers, one research engineer and six partner universities.

This research program is funded by the Riksbankens Jubileumsfond under reference number M21-0021 for a total of 33.5 Million SEK.

People and places

# Change is Key! is on Huggingface

## Models 4



Sort: Recently updated

ChangeIsKey/change-type-classifier

Text Classification • Updated 5 days ago • 11

ChangeIsKey/llama-7b-lexical-substitution

Updated Apr 21 • 16

ChangeIsKey/roberta-kubhist2

Fill-Mask • Updated Oct 26, 2023 • 9

ChangeIsKey/bert-base-swedish-cased-kubhi...

Fill-Mask • Updated Oct 10, 2023 • 1

## Datasets 2



Sort: Recently updated

ChangeIsKey/open-riksdag

Updated Sep 13, 2023

ChangeIsKey/kubhist2

Viewer • Updated Aug 10, 2023 • 291M

# Try our new model for sense relationship classification

ChangeIsKey / **change-type-classifier** like 0

Text Classification Transformers Safetensors roberta Inference Endpoints

Model card Files and versions Community Settings

## Cross-Encoder for Word-Sense Relationship Classification

This model has been trained on word sense relations extracted from WordNet.

The model can be used to detect what kind of relationships (among homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy) occur between word senses:

Given a pair of word sense definitions, predict the sense relationship (homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy).

to make someone slightly angry or upset</s><s>to talk or act in a way that makes someone lose interest

Compute

Computation time on cpu: 0.118 s



</> JSON Output

Maximize

# Try our new model for sense relationship classification

ChangeIsKey / **change-type-classifier** like 0

Text Classification Transformers Safetensors roberta Inference Endpoints

Model card Files and versions Community Settings

## Cross-Encoder for Word-Sense Relationship Classification

This model has been trained on word sense relations extracted from WordNet.

The model can be used to detect what kind of relationships (among homonymy, antonymy, hyperonymy, hyponymy, and co-hyponymy) occur between word senses:

Given a pair of word sense definitions, predict the sense relationship (homonymy, antonymy, hyperonymy, hyponymy, and co-hyponymy).

arrive at the bank of a river or the shore of a lake or sea</s><s>to reach a place, especially at the end of a journey

Compute

Computation time on cpu: 0.242 s

hyponym	0.991
co-hyponym	0.008
homonym	0.001
hypernym	0.000
antonym	0.000

<> JSON Output Maximize

# Try our new model for sense relationship classification

ChangeIsKey / **change-type-classifier** like 0

Text Classification Transformers Safetensors roberta Inference Endpoints

Model card Files and versions Community Settings

## Cross-Encoder for Word-Sense Relationship Classification

This model has been trained on word sense relations extracted from WordNet.

The model can be used to detect what kind of relationships (among homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy) occur between word senses:

Given a pair of word sense definitions, predict the sense relationship (homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy).

The set of food items that are used to make meals at home.</s><s>The flesh of an animal used as food.

Compute

Computation time on cpu: 0.235 s

hypernym	0.886
co-hyponym	0.109
hyponym	0.004
homonym	0.001
antonym	0.000

</> JSON Output Maximize

# Try our new model for sense relationship classification

ChangeIsKey / **change-type-classifier** like 0

Text Classification Transformers Safetensors roberta Inference Endpoints

Model card Files and versions Community Settings

## Cross-Encoder for Word-Sense Relationship Classification

This model has been trained on word sense relations extracted from WordNet.

The model can be used to detect what kind of relationships (among homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy) occur between word senses:

Given a pair of word sense definitions, predict the sense relationship (homonymy, antonymy, hypernymy, hyponymy, and co-hyponymy).

very poor or inferior in quality or standard; not good or well in any manner or degree</s><s>very exceptionally good or impressive, especially in a surprising or ingenious way

Compute

Computation time on cpu: 0.249 s

antonym	0.994
• homonym	0.006
• co-hyponym	0.000
• hypernym	0.000
• hyponym	0.000

# *Thank for your attention!*



Change Is Key!

<https://www.changeiskey.org/>



[pierluigi.cassotti@gu.se](mailto:pierluigi.cassotti@gu.se)