Hands-on

Pierluigi Cassotti, pierluigi.cassotti@gu.se

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

```python
from languagechange.benchmark import DWUG

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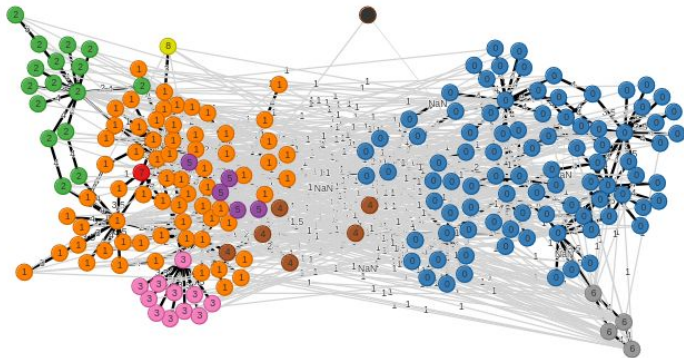
DWUGs

```python
from languagechange.benchmark import DWUG

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

dwug_en.show_usage_graph('plane_nn')
```

Computational representation of lexical items

```python
from languagechange.benchmark import DWUG
from languagechange.models.representation.contextualized import XL_LEXEME


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

```python
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)
```
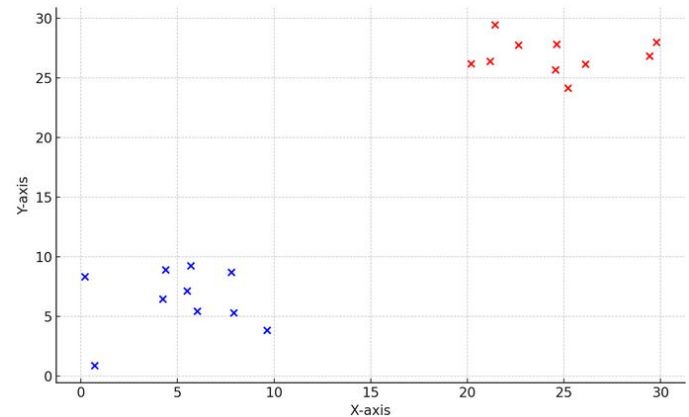
Computational representation of meaning

```
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 thi chapter, offers an interesting subject for inquiry.

T1 The we

CLUSTER 1

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.

CLUSTER 1

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

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

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

T2 The en on cells ot

T2 Lung tr days | witl

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

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.

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.benchmark import DWUG
from languagechange.models.representation.contextualized import XL_LEXEME
from languagechange.models.change.change 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)
```
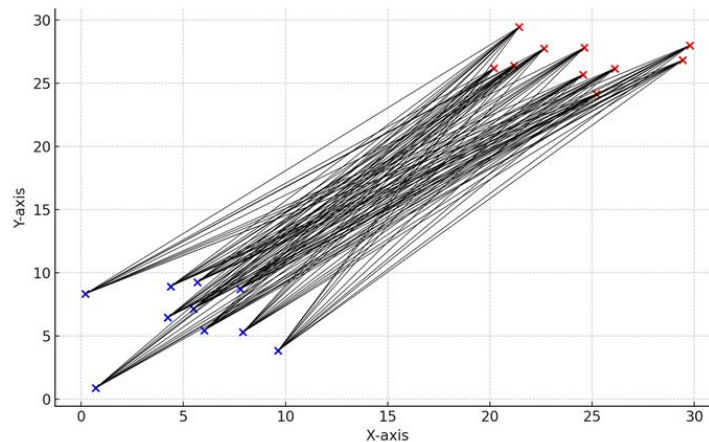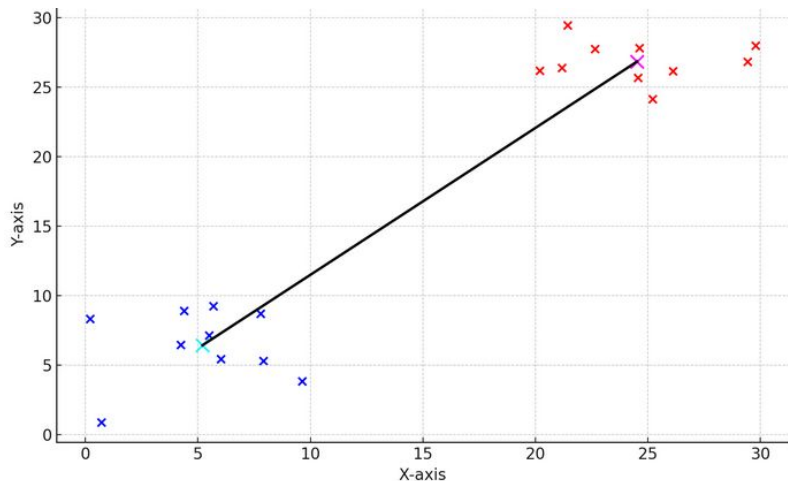
LSC metrics: PRT

```python
from languagechange.benchmark import DWUG
from languagechange.models.representation.contextualized import XL_LEXEME
from languagechange.models.change.change 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

```python
from languagechange.benchmark import DWUG
from languagechange.models.representation.contextualized import XL_LEXEME
from sklearn.cluster import AgglomerativeClustering
from languagechange.models.change.change import APD, PRT, PJSD

word = 'plane_nn'

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(clustering_algorithm=clustering)

pjsd_metric.compute_scores(vectors_time1,vectors_time2)
```
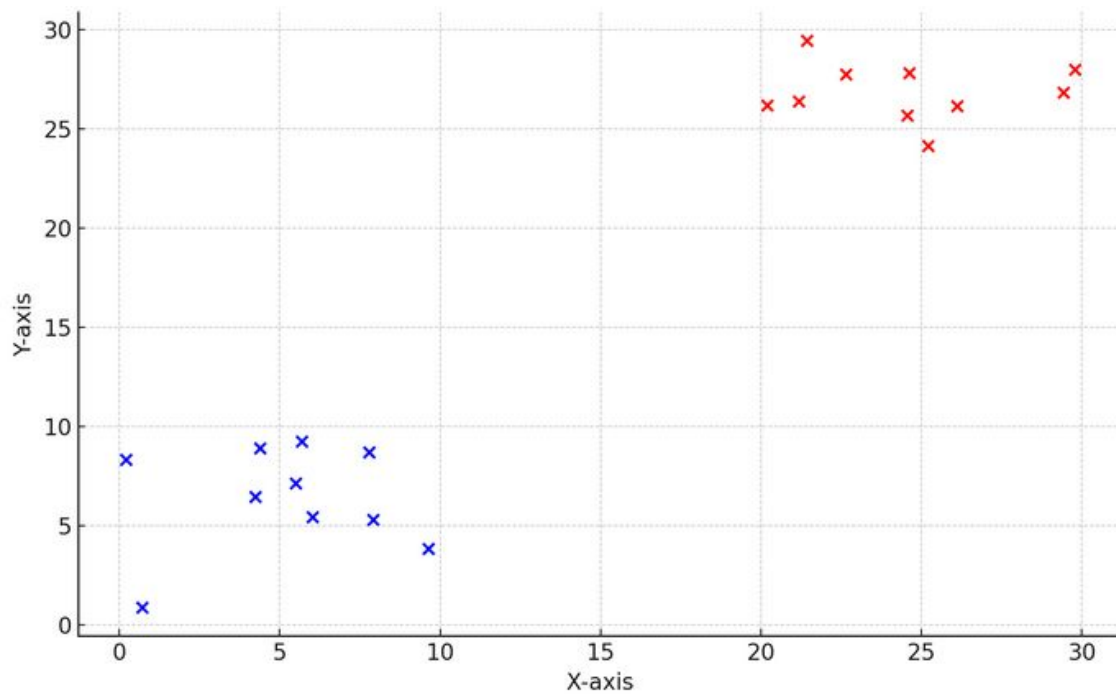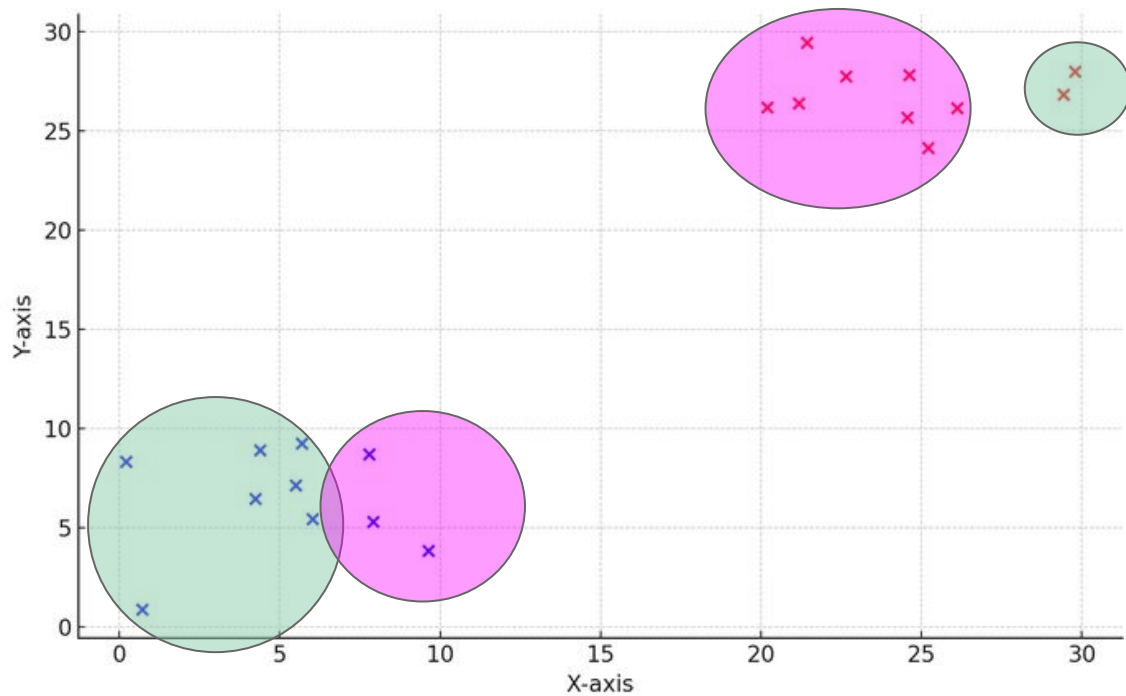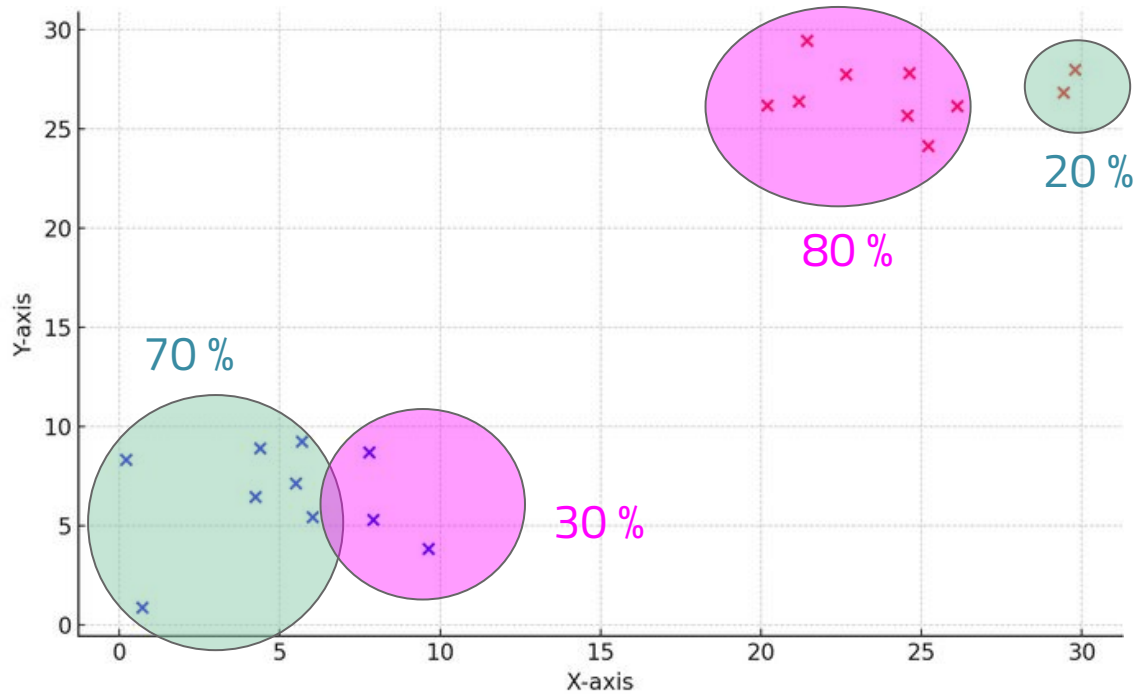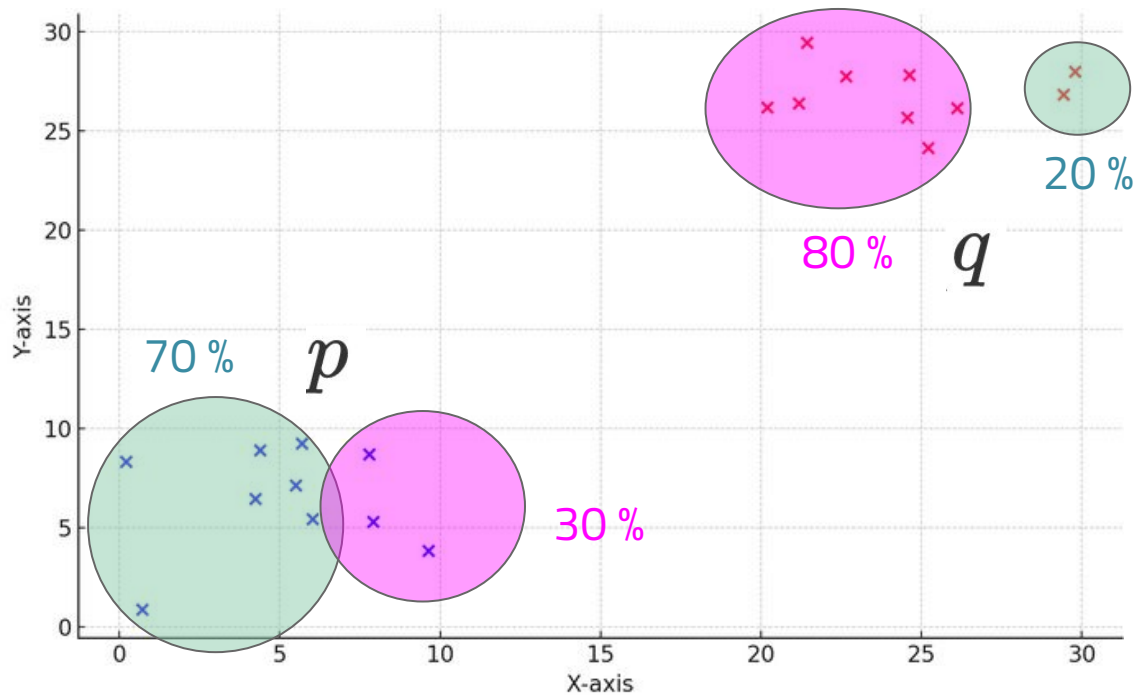
LSC metrics: PJSD

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