

Human annotation of Lexical Semantic Change

Pierluigi Cassotti, pierluigi.cassotti@gu.se

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)


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



4: Identity
3: Context variance
2: Polysemy
1: Homonymy

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 ← T2 "...overlooking and **arm** of the sea..."
"...to pay for the **arms** and ammunition..." T1 ← T2 "...an **arm** of the Red Sea..."
"...within reach of his **arm**..." T1 ← T2 "...when the disembodied **arm** of the Statue..."



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

T1

På Caudia mar fullkomlige » lugnt anda lill flutet af April, dä ön hastigt säg sig **blockerad** af 31 Cassotista och

T1

På Caudia mar fullkomlige » lugnt anda lill flutet af April, dä ön hastigt säg sig **blockerad** af 31 Cassotista och

4

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

T2

dels förtröstade man på, att folkpartiet, när det komme till **kritan**, hellre skulle taga till och med denna skenreform än — afslag.

4

T2

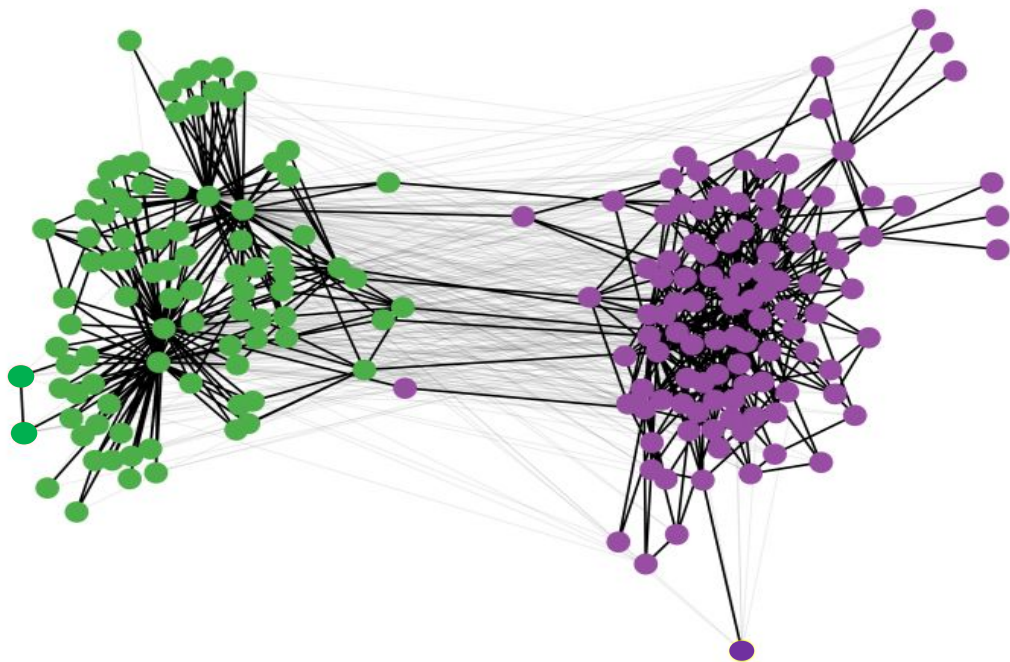
kontant efter den i lisra punkten stadgade grund. till det andra genom en oändlig rad gradationer, och alla eiler nästan alla fenomen, tilldragniog, suggestion etc., som man frambringar under sömnen, kunna hos de flesta **medier** lika val framkallas i vaket tillstånd.

T1

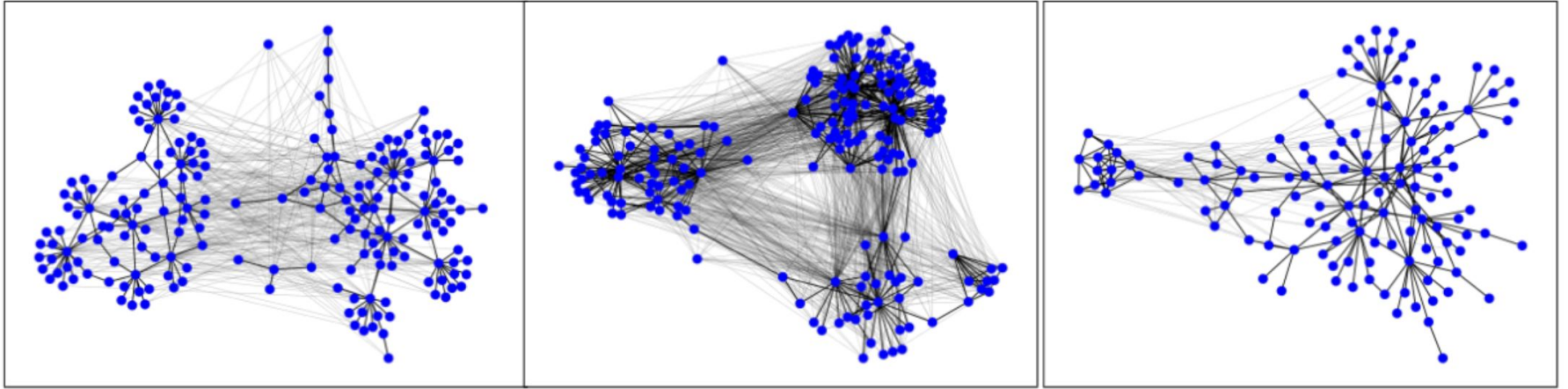
löpande årets Markegång uppgår till lika belopp eller vsverstiger det salunda uträknade **medium** as scdnaste 5 Arens markegångspris, far ingen uppsägning till lefiverering tn Nalilia ega ruin,

1

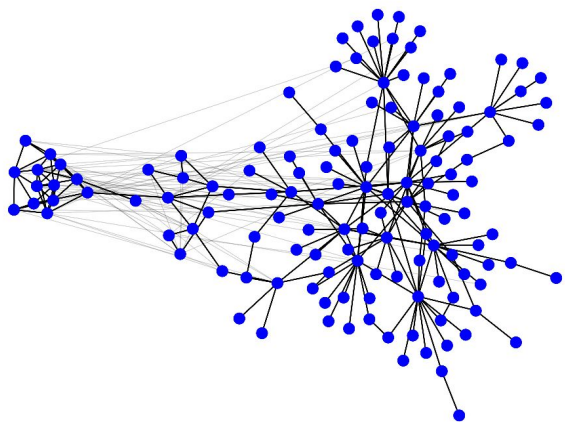
Graph view of the annotated data



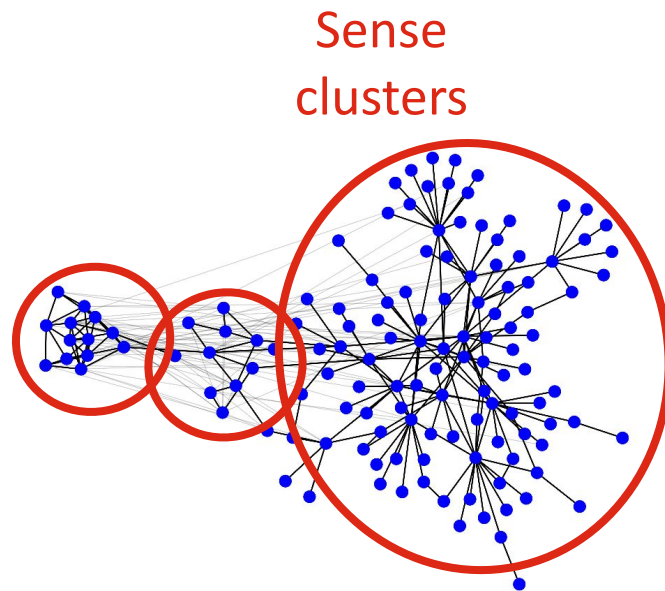
- 4: Identical
- 3: Closely related
- 2: Distantly related
- 1: Unrelated
- Nodes from time T1
- Nodes from time T2
- No Edge = No annotation



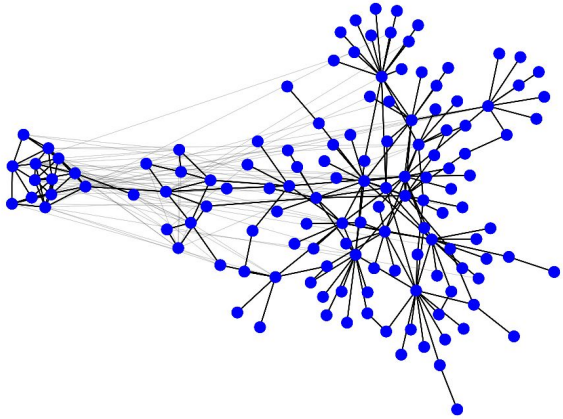
How to measure change?



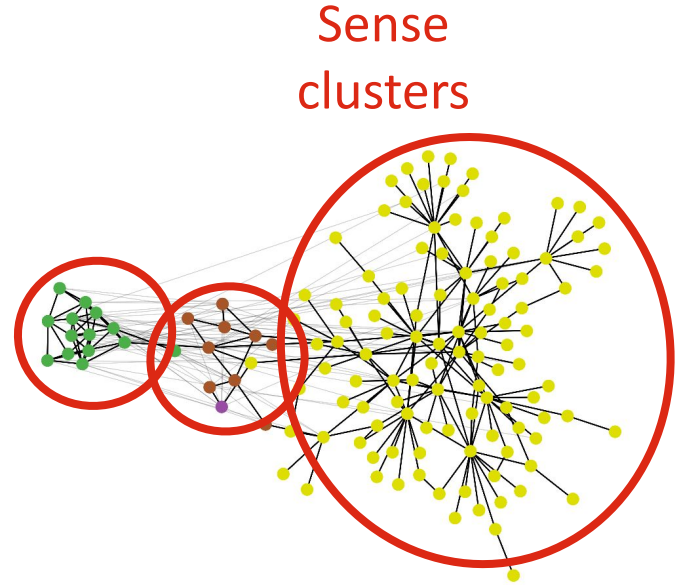
Clustering



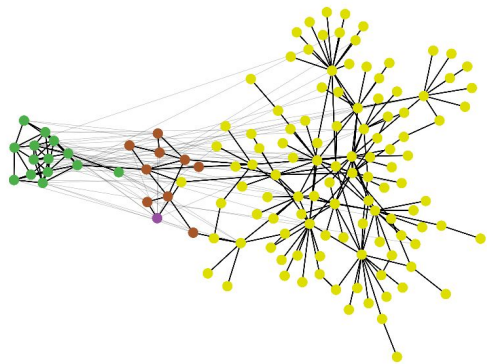
How to measure change?



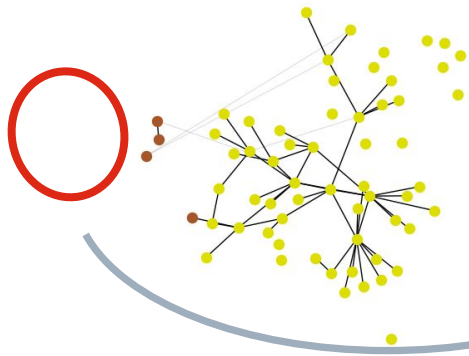
Clustering



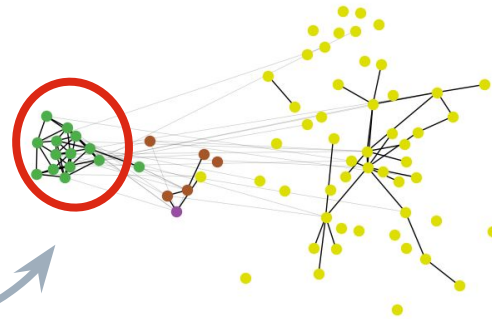
Full clustered graph



T1

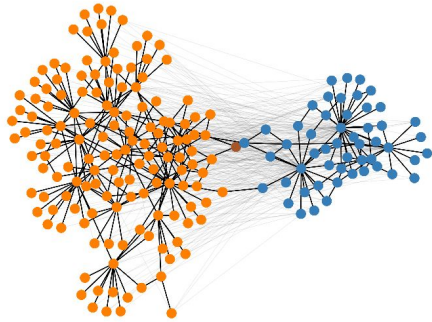


T2

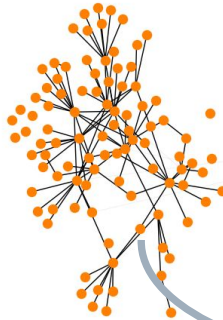


New cluster -> Binary change

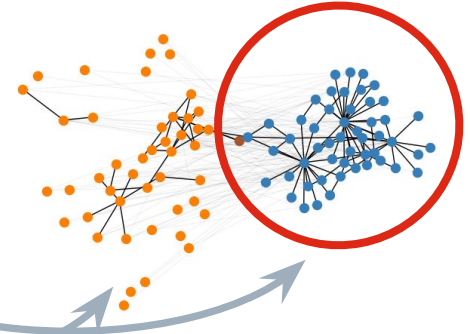
Full clustered graph



T1



T2

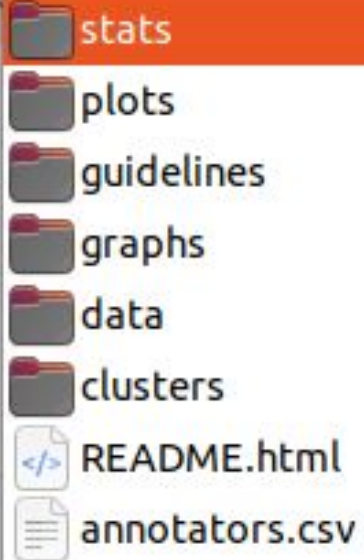


No new cluster
change in frequency of clusters
-> Graded change

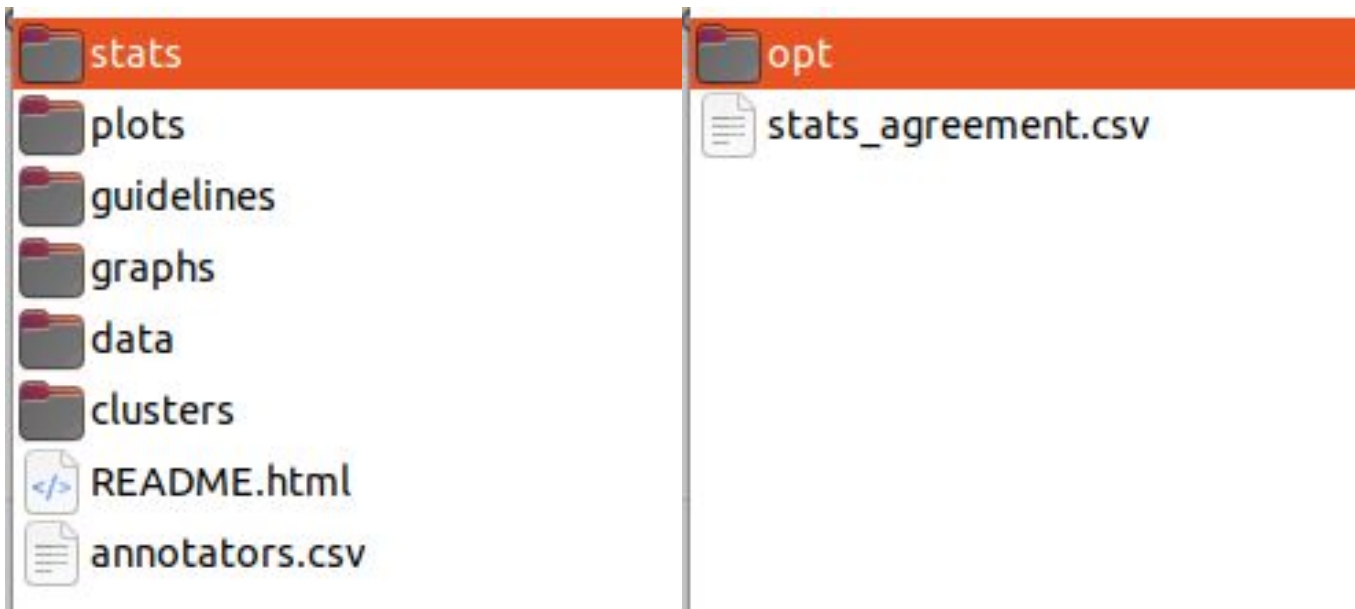
DWUGs

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

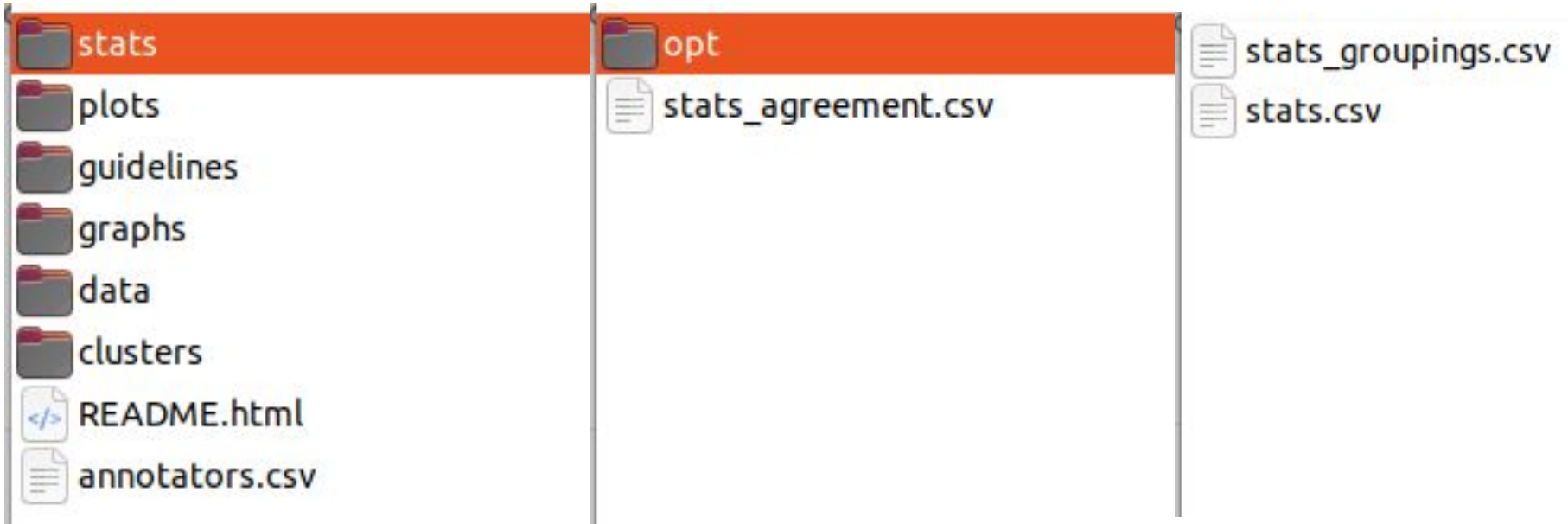
DURel Output



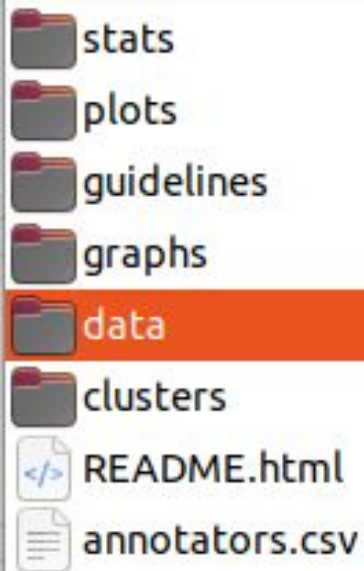
DURel Output



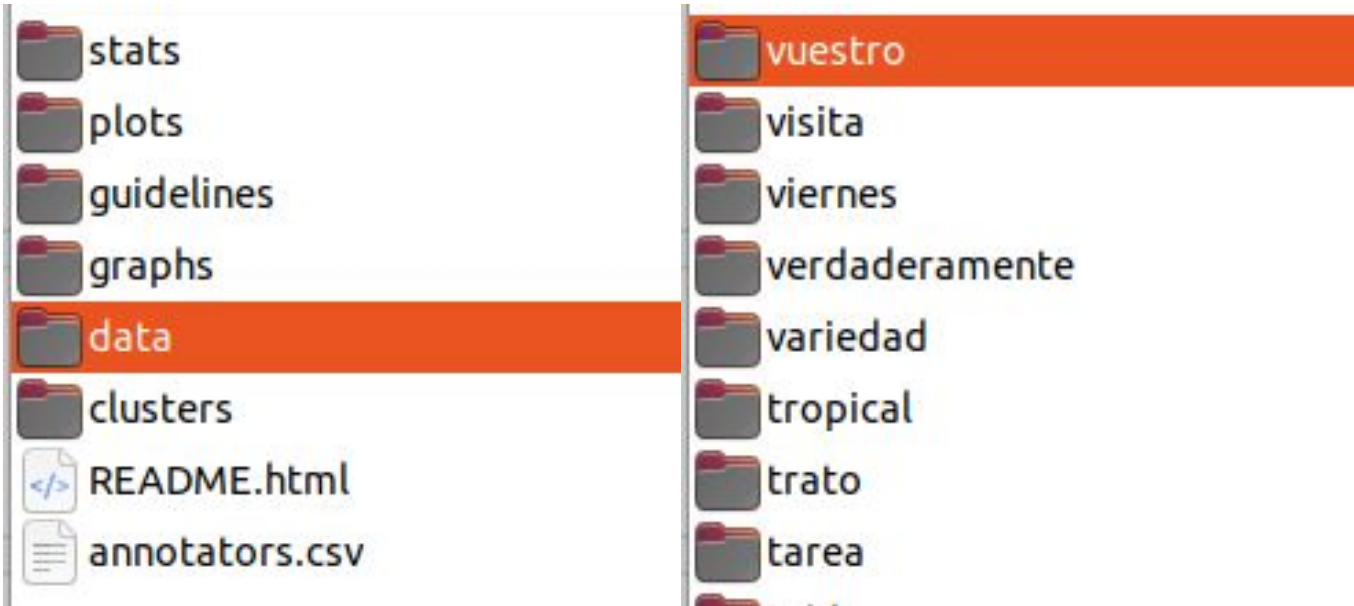
DURel Output



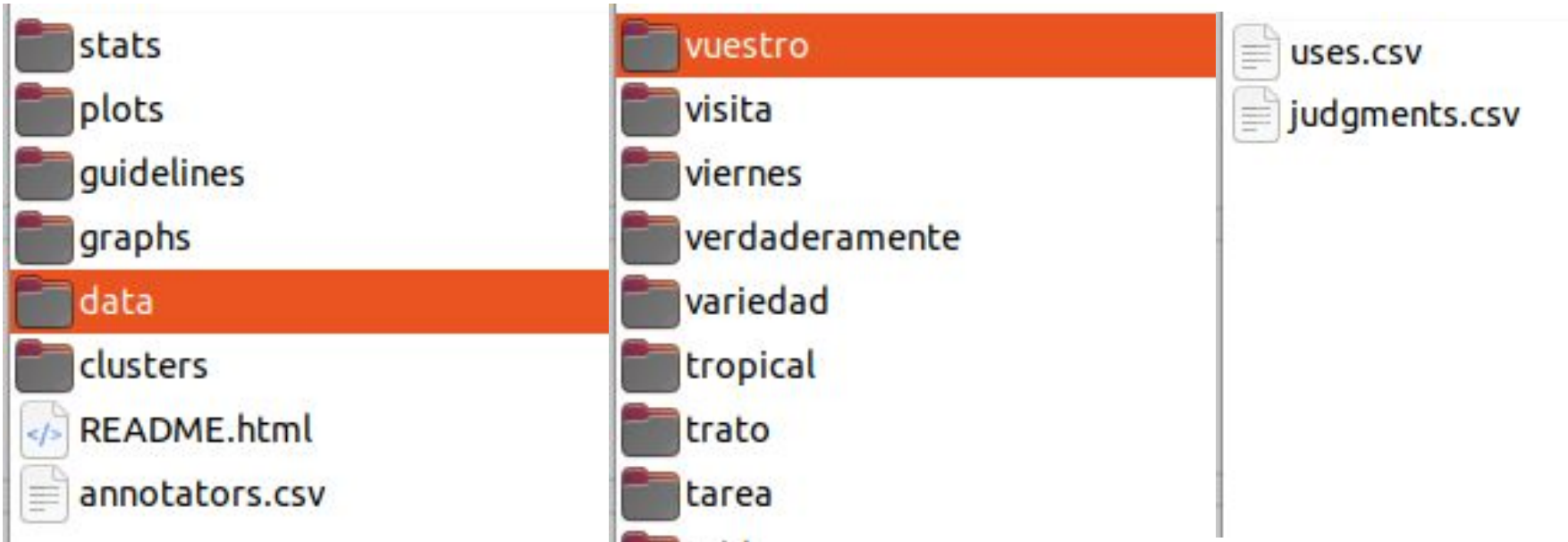
DURel Output



DURel Output



DURel Output



DURel Output

stats

plots

guidelines

graphs

data

clusters

README.html

annotators.csv

DURel Output

stats

plots

guidelines

graphs

data

clusters

README.html

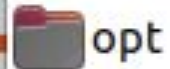
annotators.csv

opt

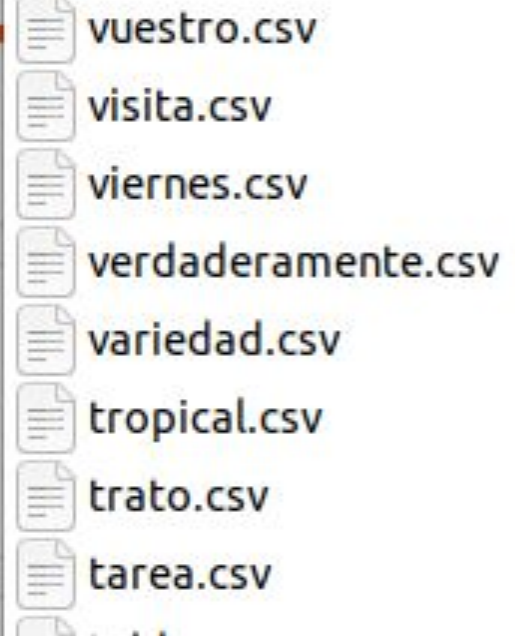
DURel Output



stats
plots
guidelines
graphs
data
clusters
README.html
annotators.csv



opt



vuestro.csv
visita.csv
viernes.csv
verdaderamente.csv
variedad.csv
tropical.csv
trato.csv
tarea.csv
...

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-116262-270	0

TRY DUREL!



[Home](#) | [Guidelines](#) | [Docs](#) | [About](#)

[Log In](#)

[Sign Up](#)



Welcome!

DUREL is a tool for annotating sentence pairs of words. Annotations of words are used to create clusters of meanings that can be visualized over time. This way, lexical properties such as polysemy, vagueness or the change in meaning of words can be examined.

We recommend Mozilla Firefox

To use this website we recommend Mozilla Firefox. Please do not use Safari, as it will cause some errors that will make it impossible to use the website correctly.

Getting Started with DUREL

If you are looking for a short and concise read about the DUREL Annotation Tool, please take a look at our [blog post](#). If you want to find out more about how to use the DUREL system, you can also check out [this video](#).

Questions and Collaboration

If you have any questions, would like to use the tool for your work or would like to collaborate with us - for example to add a new language - please do not hesitate to contact us at durel@ims.uni-stuttgart.de.

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