Computational models for Lexical Semantic Change

Francesco Periti, francesco.periti@unimi.it

CIAO!

Francesco Periti

3rd year - PhD Student

Natural Language Processing
Computational Semantics
Lexical Semantic Change

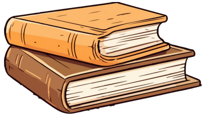
francesco.periti@unimi.it

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

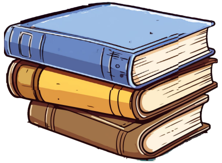
$$\mathcal{C} = \bigcup_{i=1}^{n} C_i$$

C_1



$1700 - 1800$

C_2



$1800 - 1900$

C_3

$1900 - 2024$

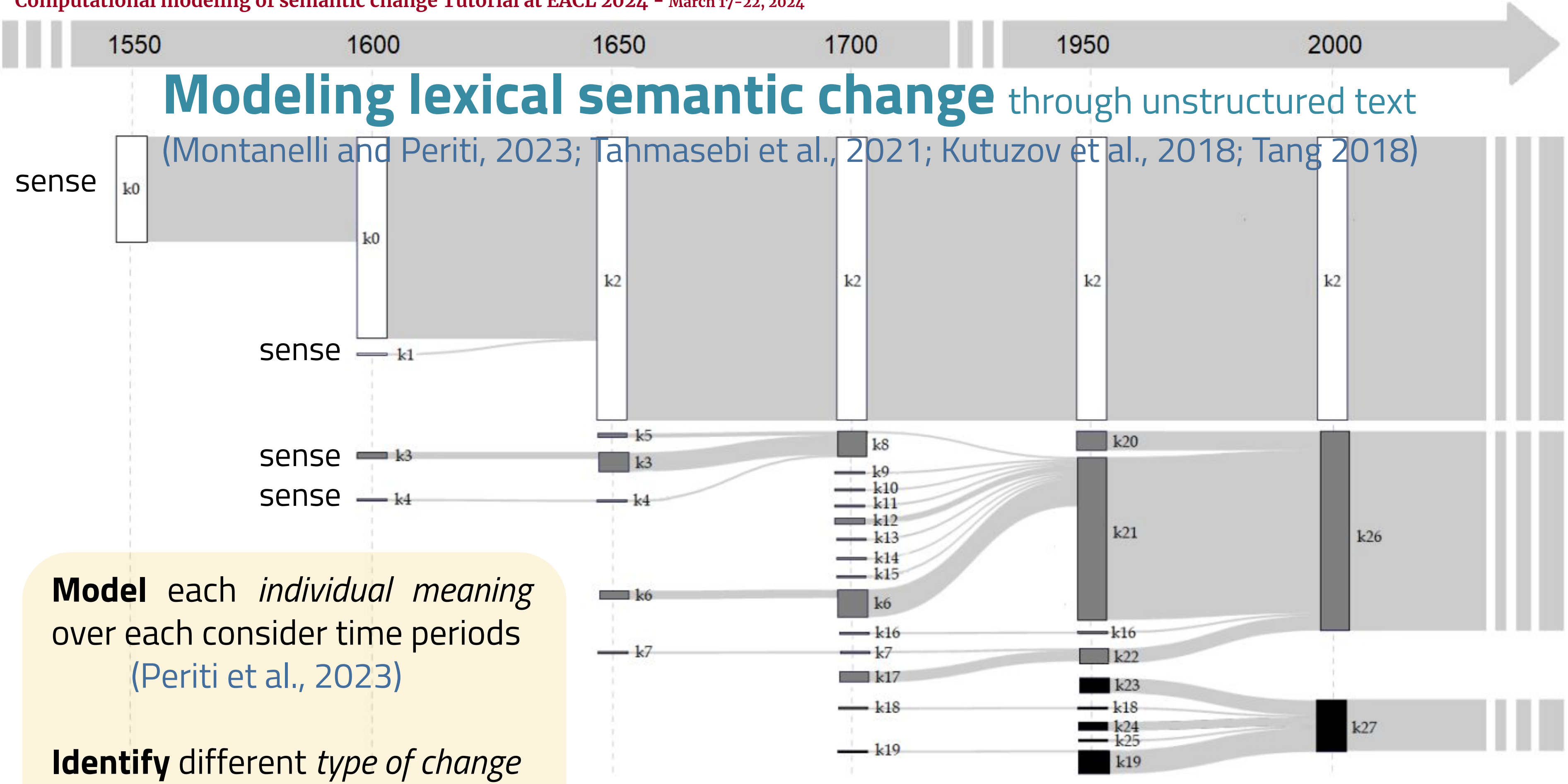
detecting
Lexical Semantic Change

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

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

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)



sense

sense

sense

sense

Model each *individual meaning* over each consider time periods (Periti et al., 2023)

Identify different *type of change* (de Sá et al., 2024)

(Castano et al., 2024)

Modeling lexical semantic change through unstructured text

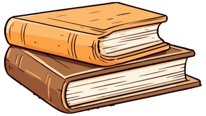
(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

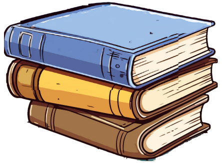
$$\mathcal{C} = \bigcup_{i=1}^{n} C_i$$

C_1



$1700 - 1800$

C_2



$1800 - 1900$

C_3

$1900 - 2024$

detecting
Lexical Semantic Change

w

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

Modeling lexical semantic change through unstructured text

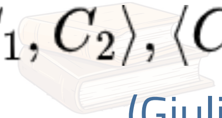
(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

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

$\langle C_1, C_2 \rangle, \langle C_2, C_3 \rangle, \ldots, \langle C_{n-1}, C_n \rangle$

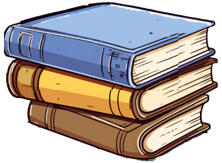
(Giulianelli et al., 2020)

C_1



$1700 - 1800$

C_2



$1800 - 1900$

C_3

$1900 - 2024$

detecting
Lexical Semantic Change

w

`gay`
From *cheerful*
To *homosexual*

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

$\langle t_1, t_2 \rangle$ Swedish (Schlechtweg et al., 2020)

$\langle t_1, t_2 \rangle$ English (Schlechtweg et al., 2020)

$\langle t_1, t_2 \rangle$ German (Schlechtweg et al., 2020)

$\langle t_1, t_2 \rangle$ Latin (Schlechtweg et al., 2020)

$\langle t_1, t_2 \rangle$ Italian (Basile et al., 2020)

$\langle t_1, t_2 \rangle$ Spanish (Zamora-Reina et al., 2022)

$\langle t_1, t_2 \rangle$ Norwegian (Kutuzov et al., 2022)

$\langle t_1, t_2 \rangle$ Chinese (Chen et al., 2023)

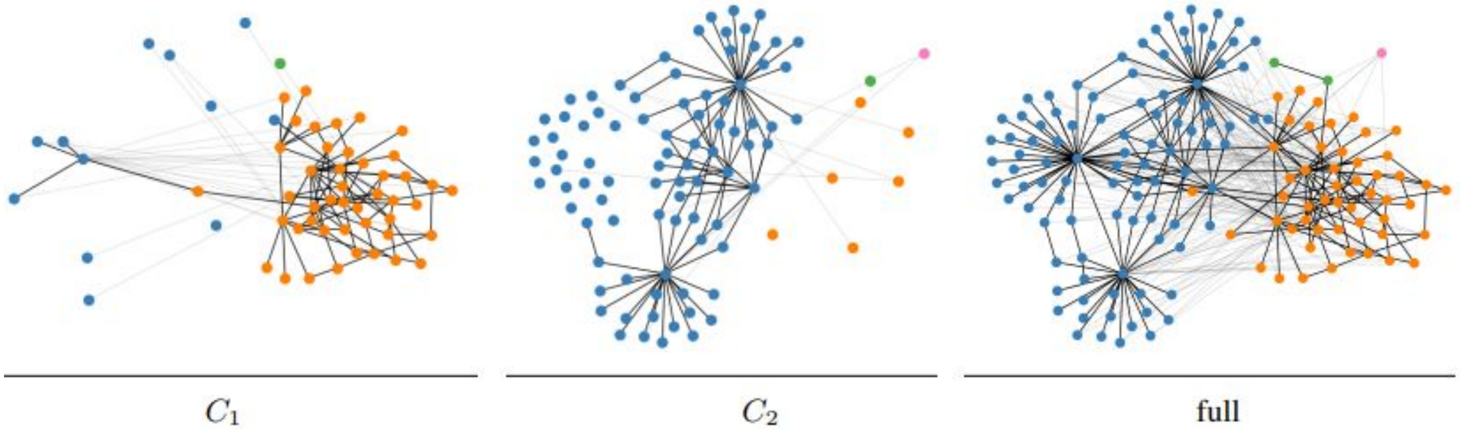
$\langle t_1, t_2 \rangle$ Japanese (Ling et al., 2023)

$\langle t_1, t_2 \rangle$ Slovenian (Pranjić et al., 2024)

$\langle t_1, t_2, t_3 \rangle$ Russian (Kutuzov and Pivovarova., 2021)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)



Human annotators

The adopted **DWUG** paradigm requires extensive annotation efforts

(Schlechtweg et al., 2021)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

44 Swedish (Schlechtweg et al., 2020)

46 English (Schlechtweg et al., 2020)

50 German (Schlechtweg et al., 2020)

40 Latin (Schlechtweg et al., 2020)

18 Italian (Basile et al., 2020)

100 Spanish (Zamora-Reina et al., 2022)

40 Norwegian (Kutuzov et al., 2022)

40 Chinese (Chen et al., 2023)

9 Japanese (Ling et al., 2023)

105 Slovenian (Pranjić et al., 2024)

111 Russian (Kutuzov and Pivovarova., 2021)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

graded binary — Swedish (Schlechtweg et al., 2020)

graded binary — English (Schlechtweg et al., 2020)

graded binary ***sense*** — German (Schlechtweg et al., 2020)

graded binary — Latin (Schlechtweg et al., 2020)

binary — Italian (Basile et al., 2020)

graded binary — Spanish (Zamora-Reina et al., 2022)

graded binary — Norwegian (Kutuzov et al., 2022)

graded binary — Chinese (Chen et al., 2023)

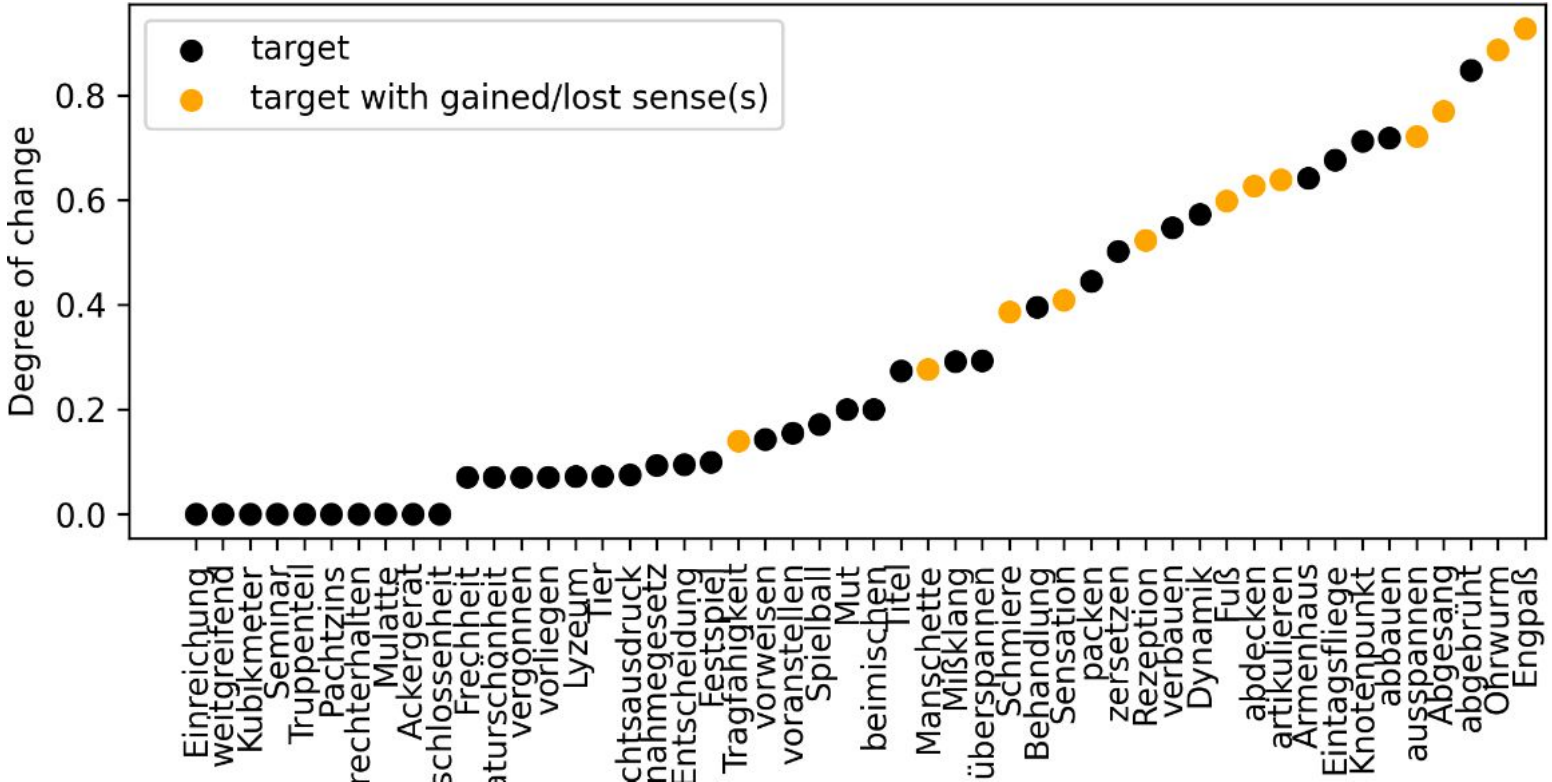
graded binary — Japanese (Ling et al., 2023)

graded — Slovenian (Pranjić et al., 2024)

graded — Russian (Kutuzov and Pivovarova., 2021)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)



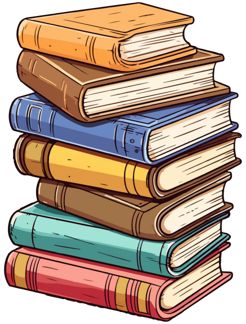
Modeling lexical semantic change through **contextualized LMs**

(Montanelli and Periti, 2023)

Word usage extraction

- If a particle of mass *m* is placed on a smooth inclined **plane** and re- leased, it will slide down the slope.
- [...]

target w : plane



Embedding generation

The **plane** flew above the clouds

Meaning representation

sub-corpus C_1

Φ_1

sub-corpus C_2

Φ_2

Shift assessment

$f(\Phi_1, \Phi_2)$

Modeling lexical semantic change through contextualized LMs
(Montanelli and Periti, 2023)

Meaning representation

form-based

sense-based

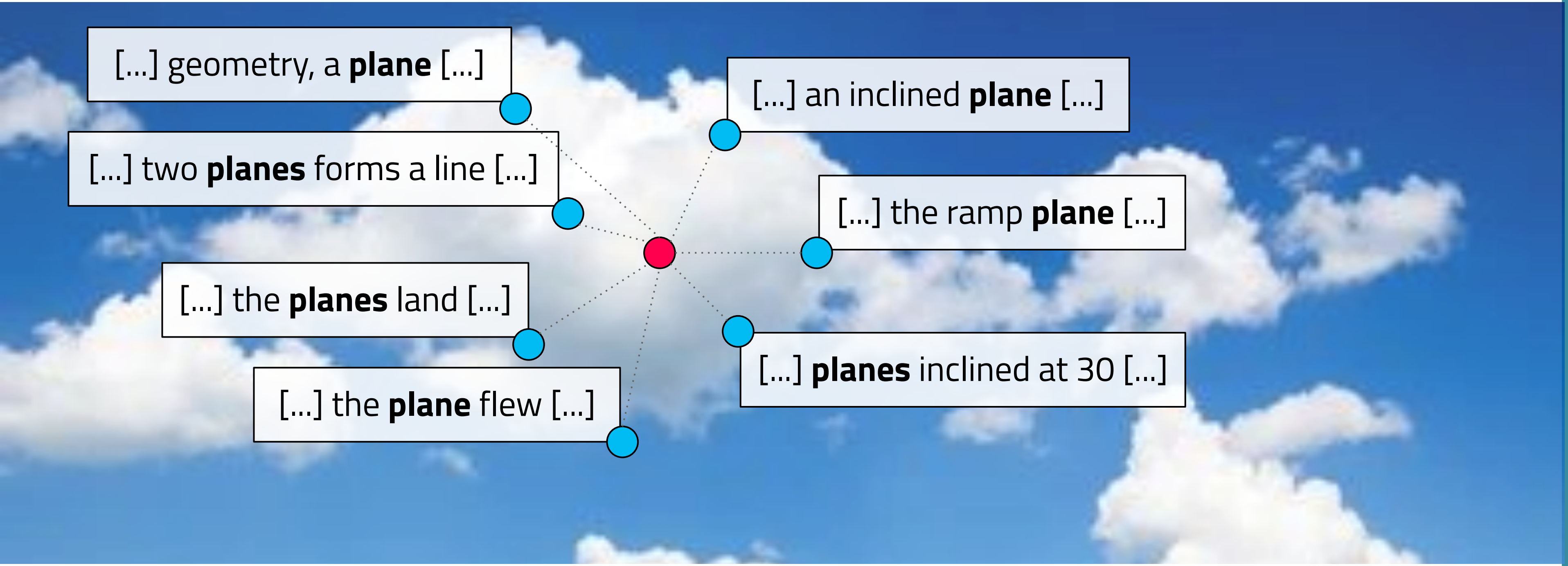
dominant meaning

polysemous structure

individual meanings

Modeling lexical semantic change through **form-based approaches**

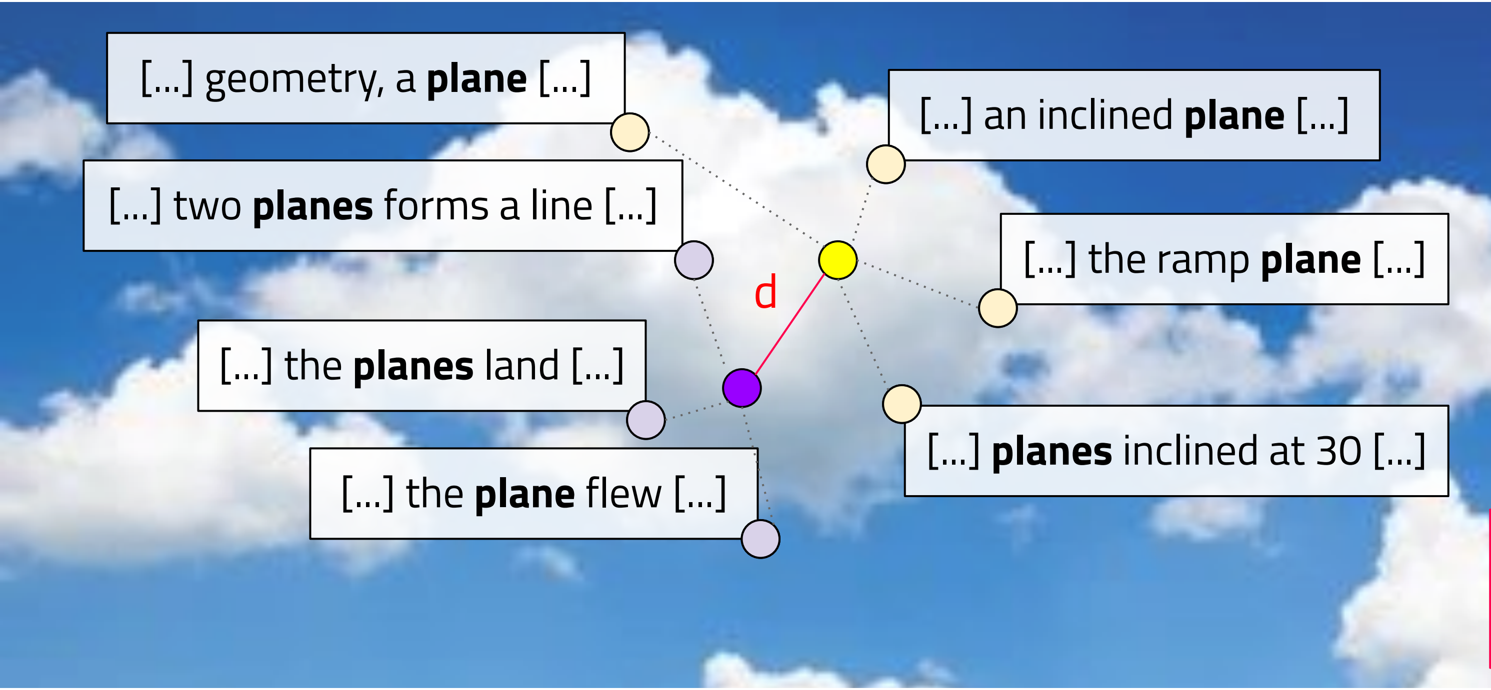
(Montanelli and Periti, 2023)



word prototype

Modeling lexical semantic change through **form-based approaches**

(Montanelli and Periti, 2023)

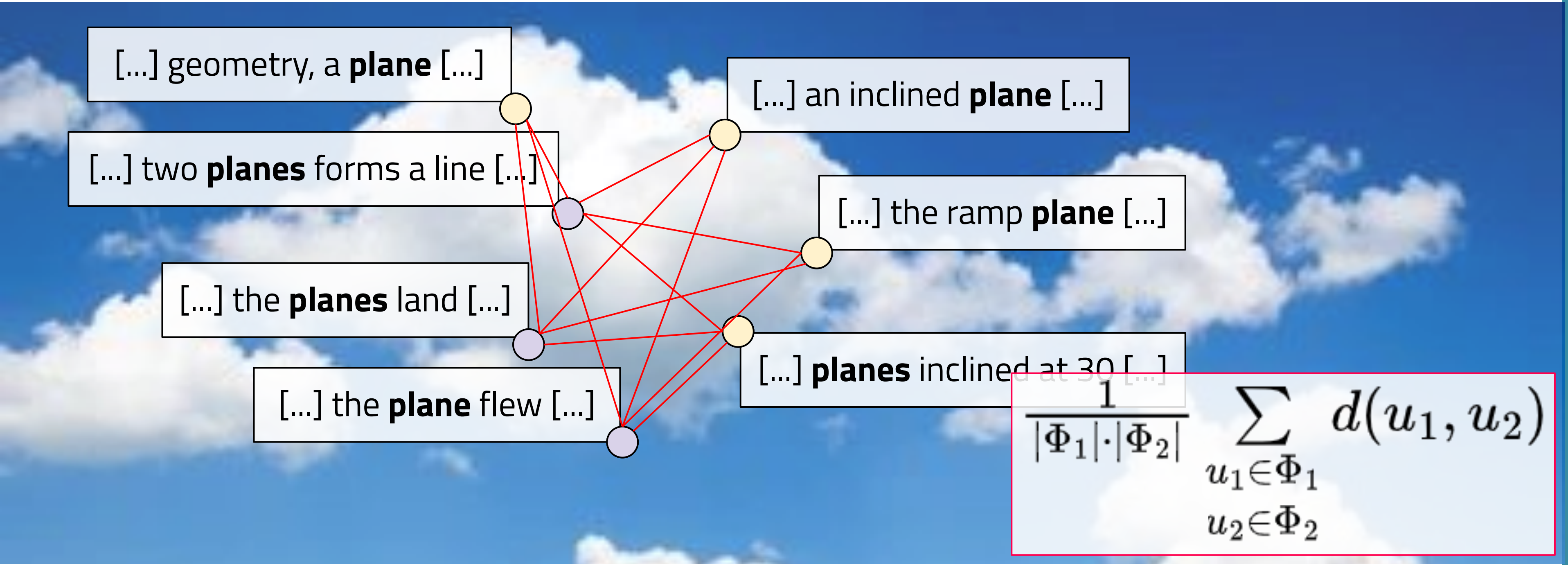


$$1 - \frac{\mu_1 \cdot \mu_2}{||\mu_1|| \, ||\mu_2||}$$

cosine distance between word PRoTotypes **(PRT)** (Martinc et al., 2020a; Kutuzov and Giulianelli, 2020)

Modeling lexical semantic change through **form-based approaches**

(Montanelli and Periti, 2023)



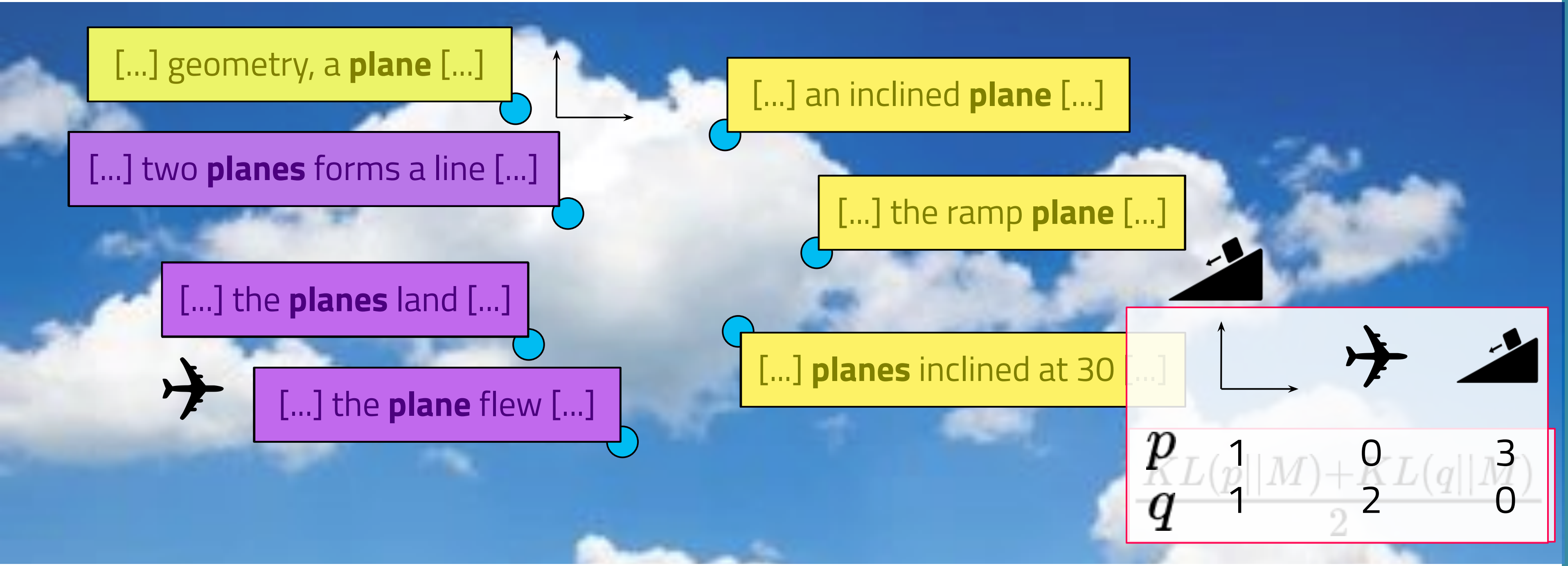
$$\frac{1}{|\Phi_1|\cdot|\Phi_2|}\sum_{\substack{u_1\in\Phi_1\\ u_2\in\Phi_2}} d(u_1,u_2)$$

Average Pairwise Distance **(APD)** (Giulianelli et al., 2020)

Modeling lexical semantic change through sense-based approaches

(Montanelli and Periti, 2023)

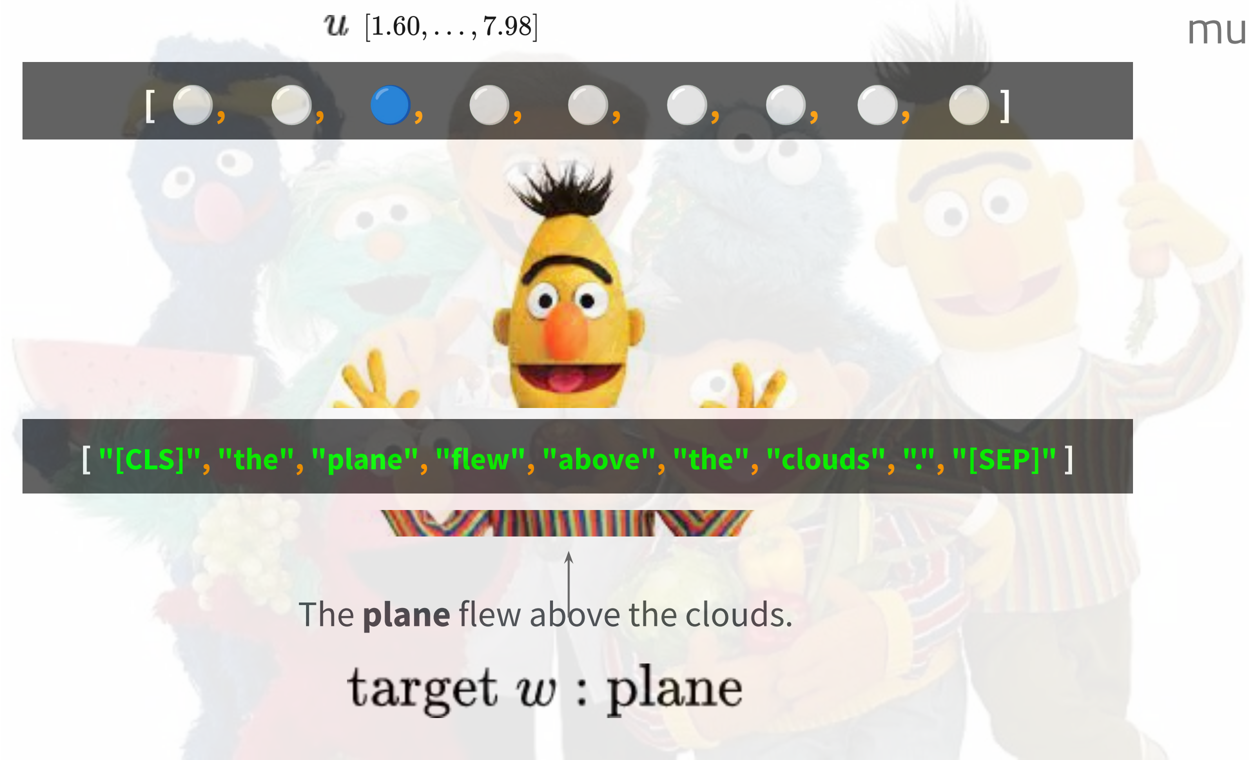
$$\text{clustering}(\Phi_1 \cup \Phi_2)$$



Affinity Propagation + Jensen Shannon Divergence **(AP+JSD)** (Martinc et al., 2020b)

Scalability issues arising from **contextualized LMs**

Memory consumption



u $[1.60, \ldots, 7.98]$

multidimensional continuous vector (i.e., embeddings)

An embedding contains **768 floats**

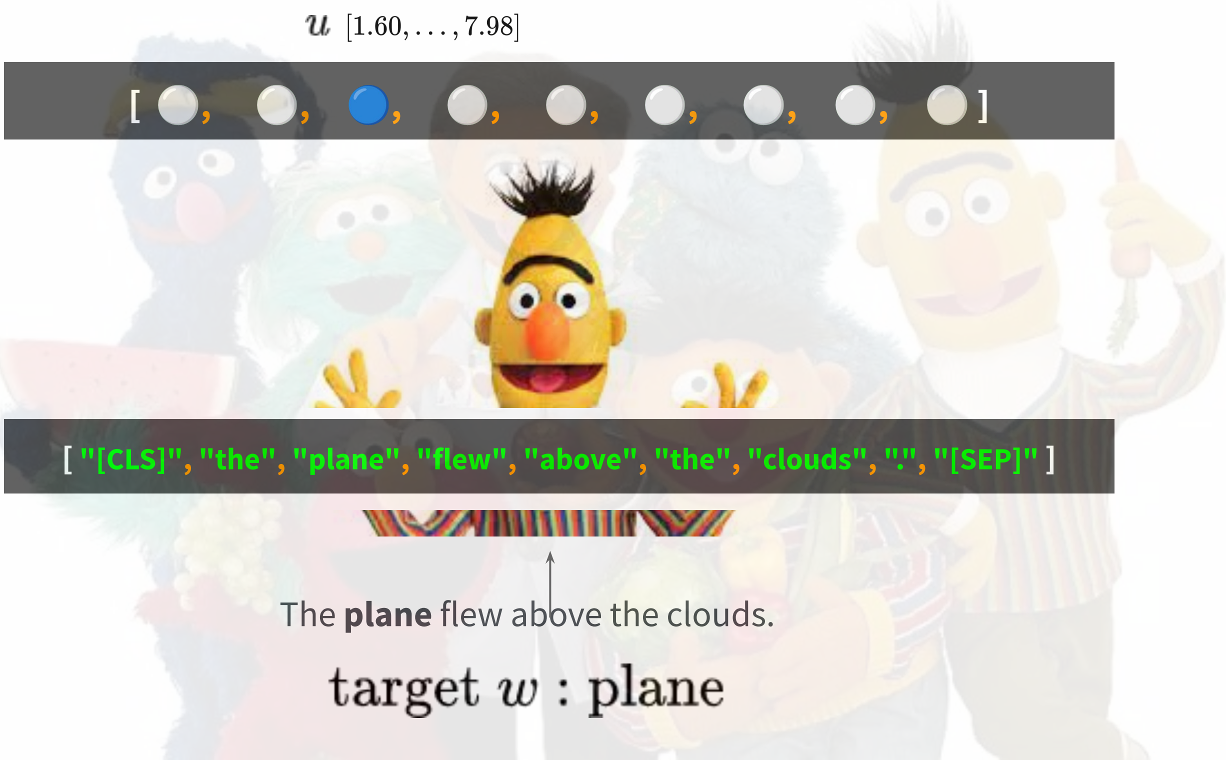
A floating-point in Python requires **8 B**

A word appearing 500.000 times requires **3Gb** + *overhead*

(Montariol et al., 2020)
(Montanelli and Periti, 2023)

Scalability issues arising from contextualized LMs

Memory consumption



Workaround

Use smaller models
(Rosin and Radinsky, 2022)

Process one target word at a time

Random sampling the occurrences
(Rodina et al., 2021)

Dimensionality reduction of the embeddings
(Rother et al., 2020)

(Montariol et al., 2020)
(Hengchen et al., 2021)
(Montanelli and Periti, 2023)

Scalability issues arising from **contextualized LMs**

Memory consumption

elicopter – sun – plane – bird – moon – ...

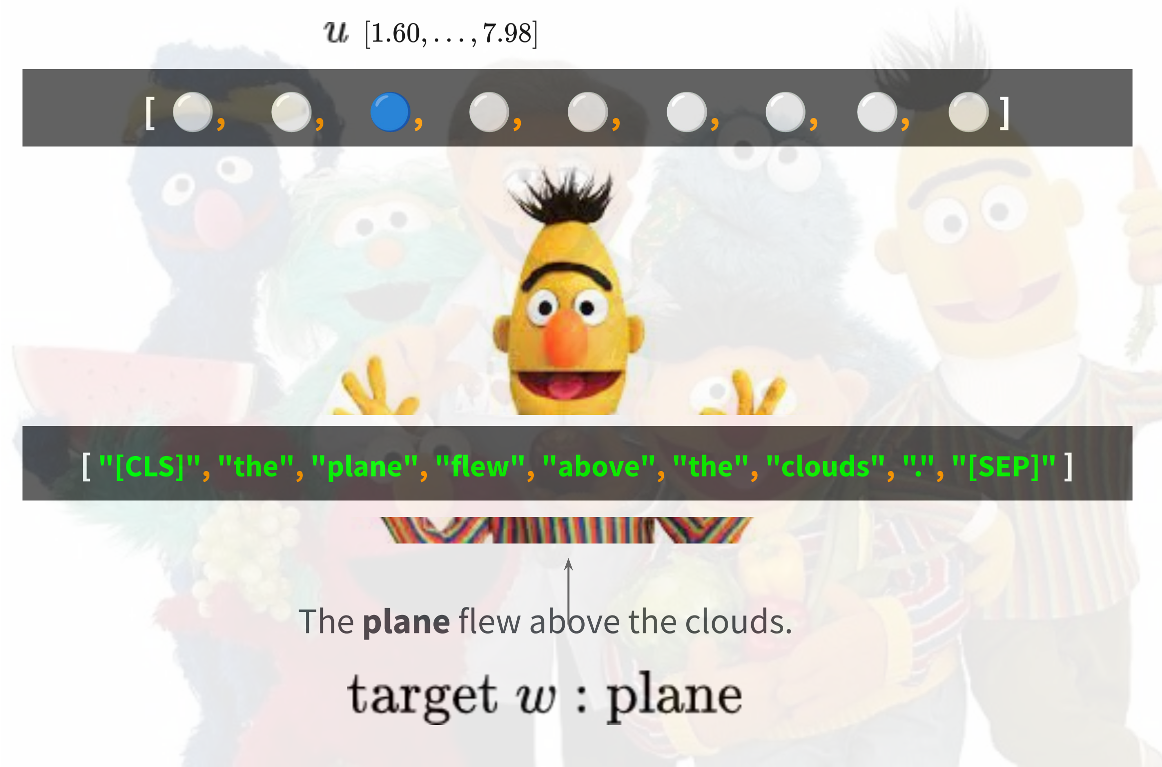
The **[MASK]** flew above the clouds.
The **plane** flew above the clouds.

Lexical Substitutes for LSC

(Arefyev and Zhikov, 2020)
(Kudisov and Arefyev, 2022)
(Card, 2023)
(Cuscito et al., 2024)

Scalability issues arising from **contextualized LMs**

Computation time



Word usage extraction

Embedding generation

Dimensionality Reduction

Embedding distance

Clustering

Typical Requirements

GPUs are required

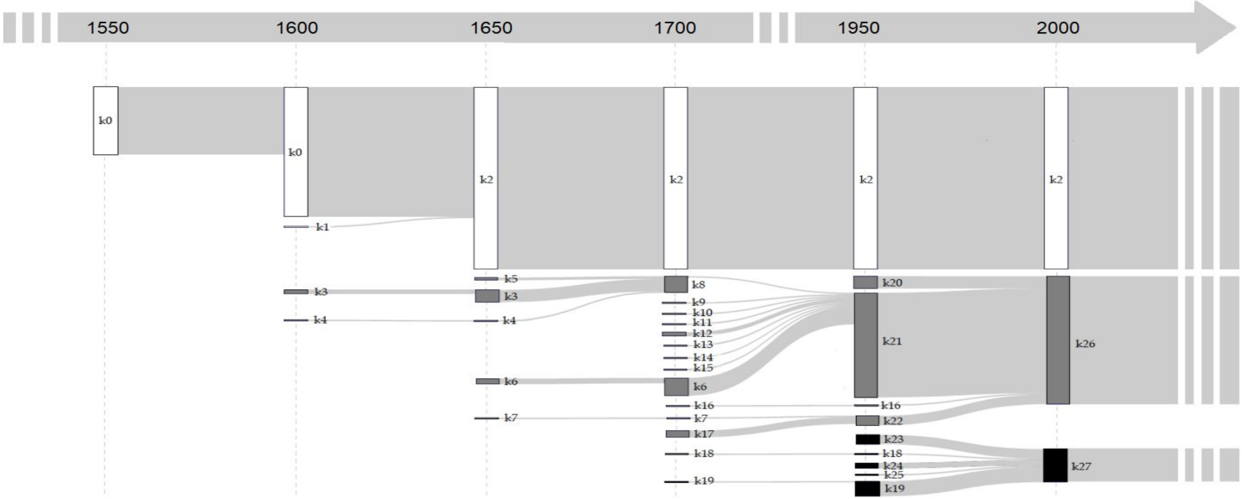
Select a small set of target words

Random sampling the occurrences

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

Where are we now?



(Castano et al., 2024)

Modeling lexical semantic change through contextualized LMs

(Montanelli and Periti, 2023)

Graded Change Detection - Spearman's rank correlation coefficient

	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 (Giulianelli et al., 2019)	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 (Martinc et al., 2020a)	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 (Martinc et al., 2020b)	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 (Periti et al., 2022)	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**

(Periti and Tahmasebi, 2024)

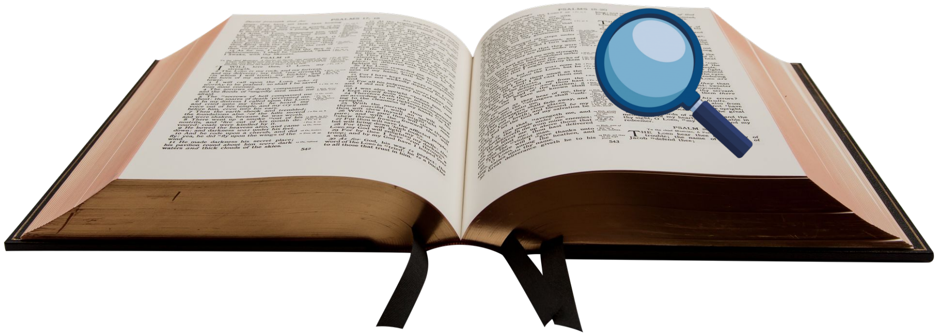
Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

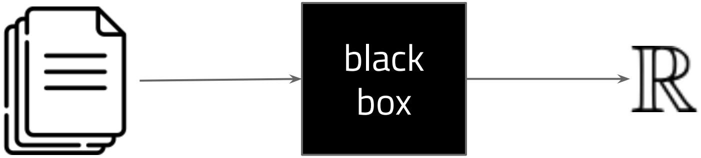
To model *lexical semantic change*, we must go beyond merely quantifying the change

Interpretability issues arising from **contextualized LMs** form-based approaches

close reading



How can they be useful?



form-based approaches are not interpretable

To identify words that have changed across multiple time periods, necessitating further sense-based modeling.

To quantify the changes at the vocabulary level.

Interpretability issues arising from contextualized LMs
sense-based approaches

Expectation

p	1	0	3
q	1	2	0

Reality

p	4	0	4	4	4
q	3	2	3	0	5

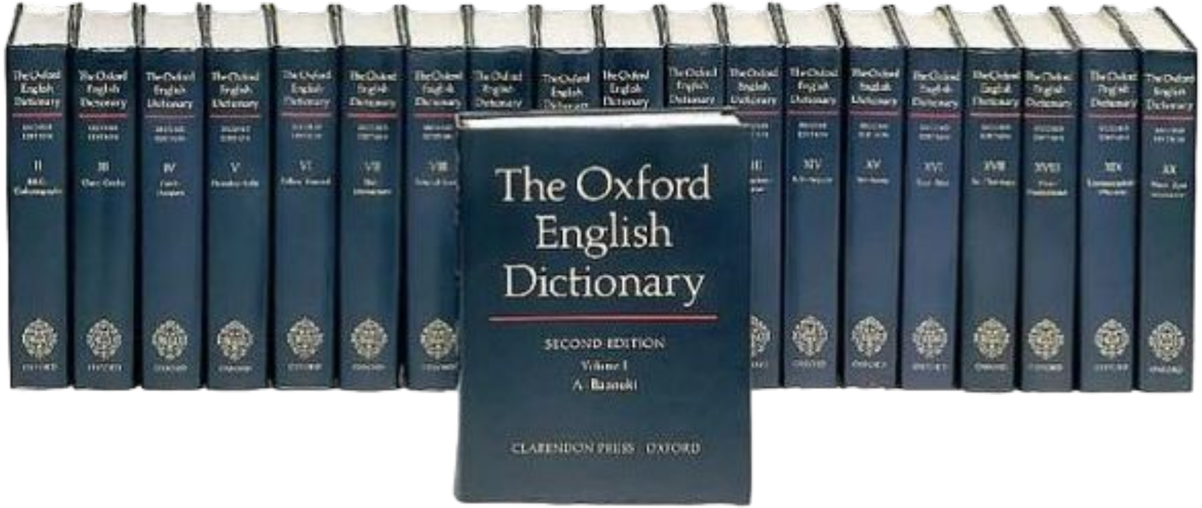
Clusters of embeddings loosely correspond to word senses
(Montanelli and Periti, 2023)
(Kutuzov et al., 2022)

LMs derive signals of meaning from contexts surrounding word tokens

Cluster of contextual usages rather than cluster of word meanings

A word may change its context without changing its meaning

To supervise or not to supervise: that is NOT the question



One one hand

Supervised models are more powerful

One the other hand

Diachronic resources are typically not available





Interpretability issues arising from **contextualized LMs**
word meaning description

	???	???	???	???	???
p	4	0	4	4	4
q	3	2	3	0	5

Do these techniques work well?

The **plane** flew above the clouds.
The **[MASK]** flew above the clouds.

elicopter – sun – bird – moon

How can we facilitate the interpretation?

Keyword extraction
(Montariol et al., 2021)

Lexical substitutes
(Kudisov and Arefyev, 2022)

Sentence Example Extraction
(Giulianelli et al., 2019)

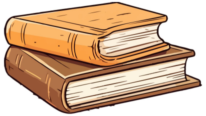
Definition generation
(Giulianelli et al., 2023)

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

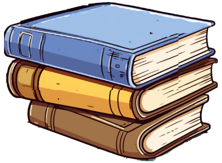
$$\mathcal{C} = \bigcup_{i=1}^{n} C_i$$

C_1



1700 – 1800

C_2



1800 – 1900

C_3

1900 – 2024

detecting

Lexical Semantic Change

w `manufacture`

From *to make by hand*

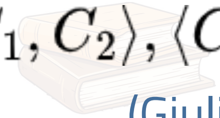
To *to make by machine*

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

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

C_1

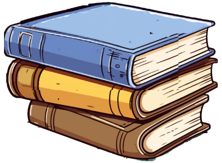


1700 – 1800

$$\langle C_1, C_2 \rangle, \langle C_2, C_3 \rangle, \ldots, \langle C_{n-1}, C_n \rangle$$

(Giulianelli et al., 2020)

C_2



1800 – 1900

C_3

1900 – 2024

detecting
Lexical Semantic Change

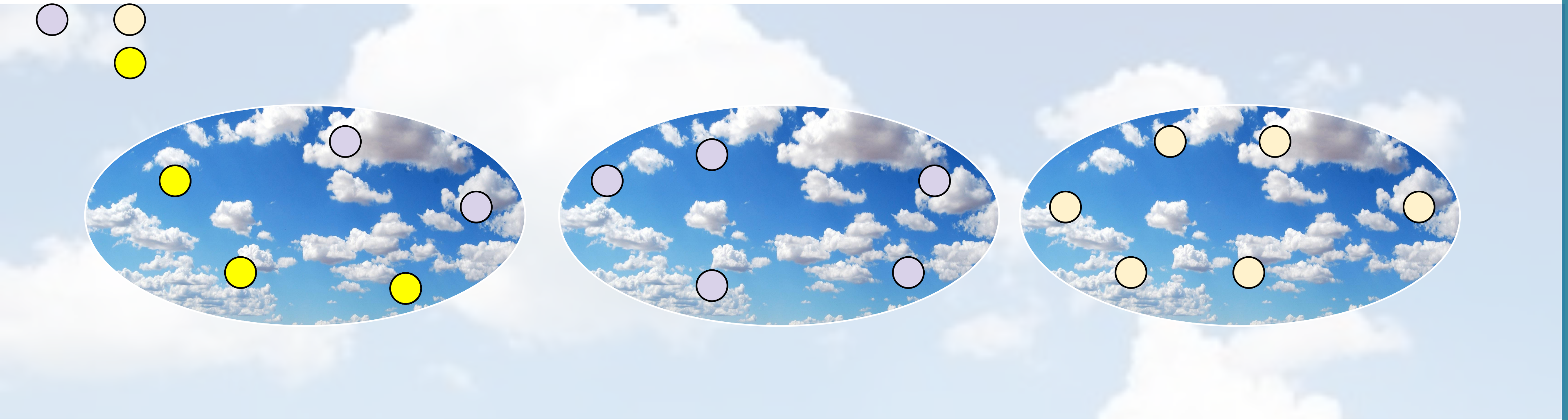
w `gay`
From *cheerful*
To *homosexual*

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution

Embedding space

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution

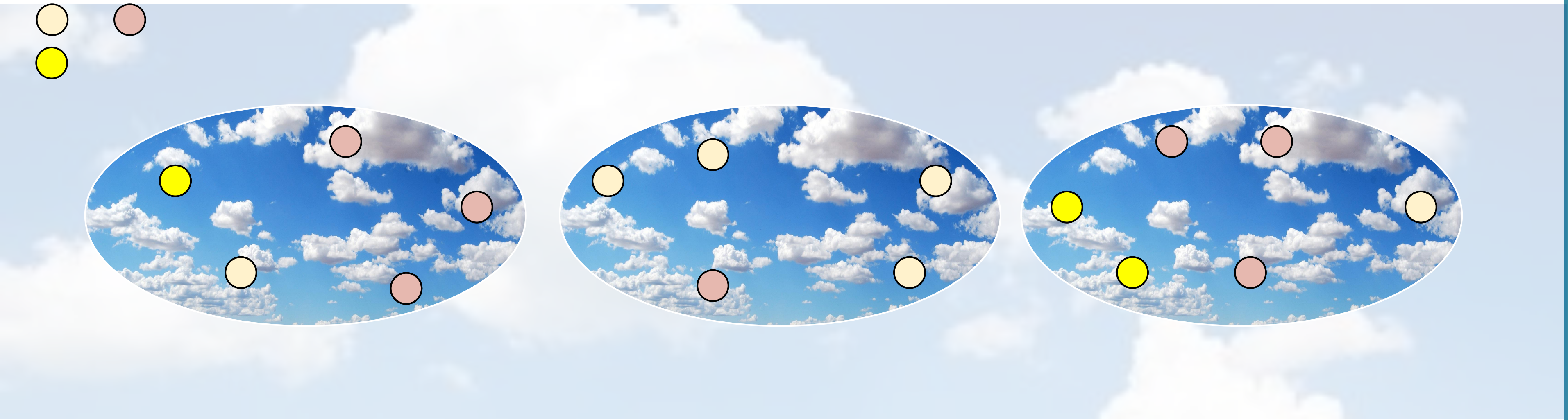
$\langle t_1, t_2 \rangle$



Clustering is typically performed jointly (Martinc et al., 2020b)

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution

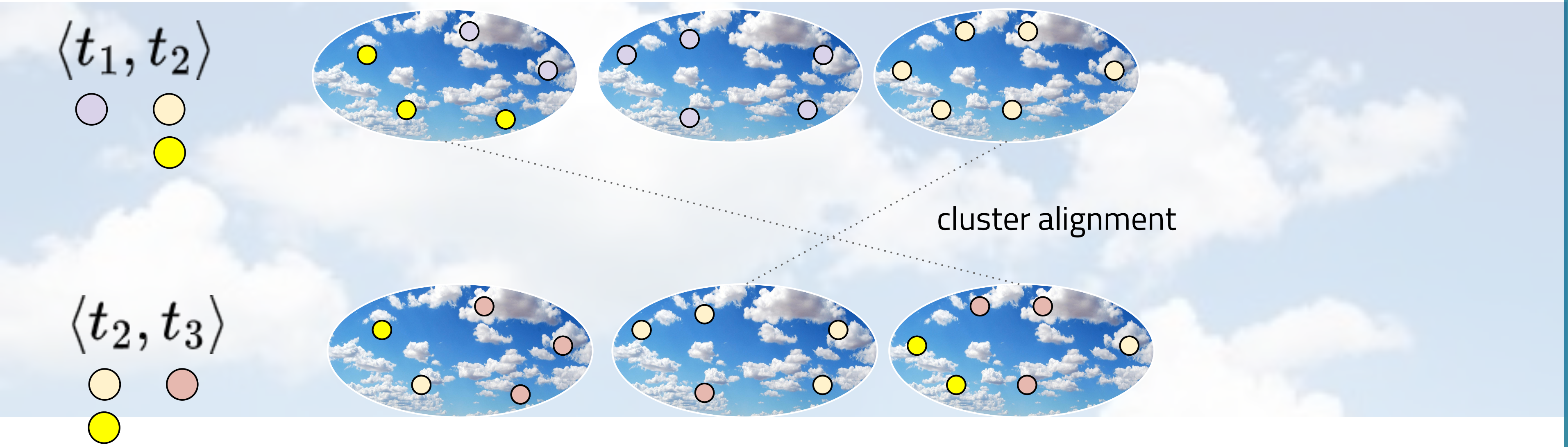
$\langle t_2, t_3 \rangle$



Clustering is typically performed jointly (Martinc et al., 2020b)

Interpretability issues arising from **contextualized LMs**

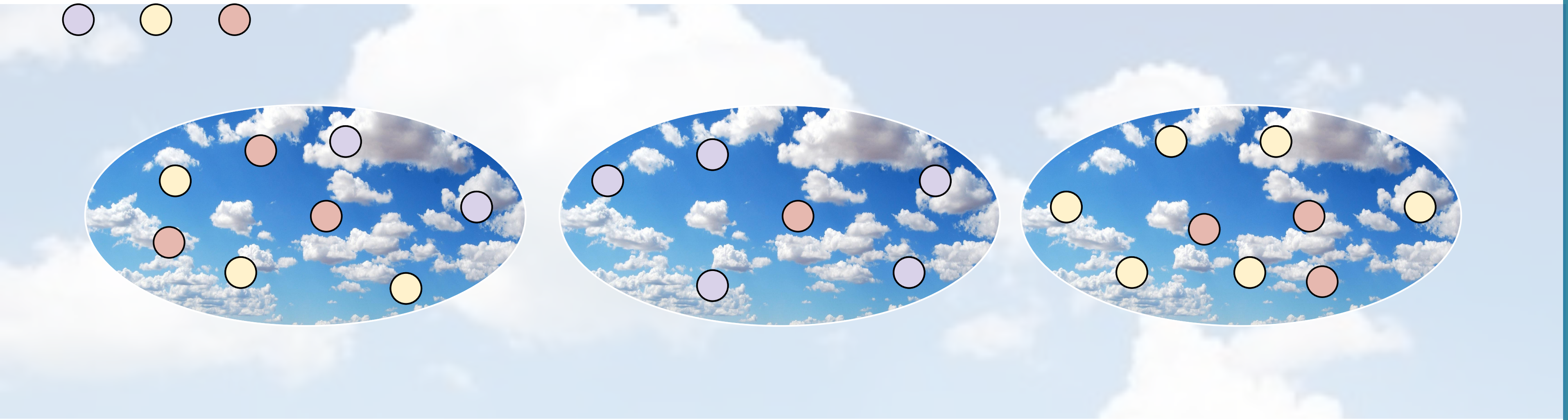
word meaning interpretation and evolution



Clustering is typically performed jointly

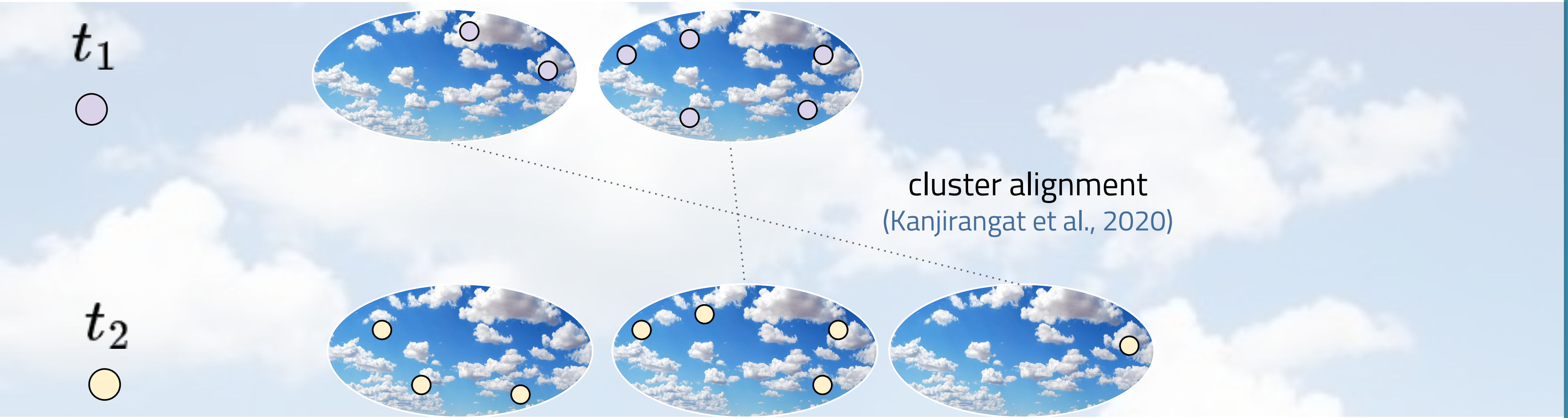
Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution

$$\langle t_1, t_2, t_3 \rangle$$



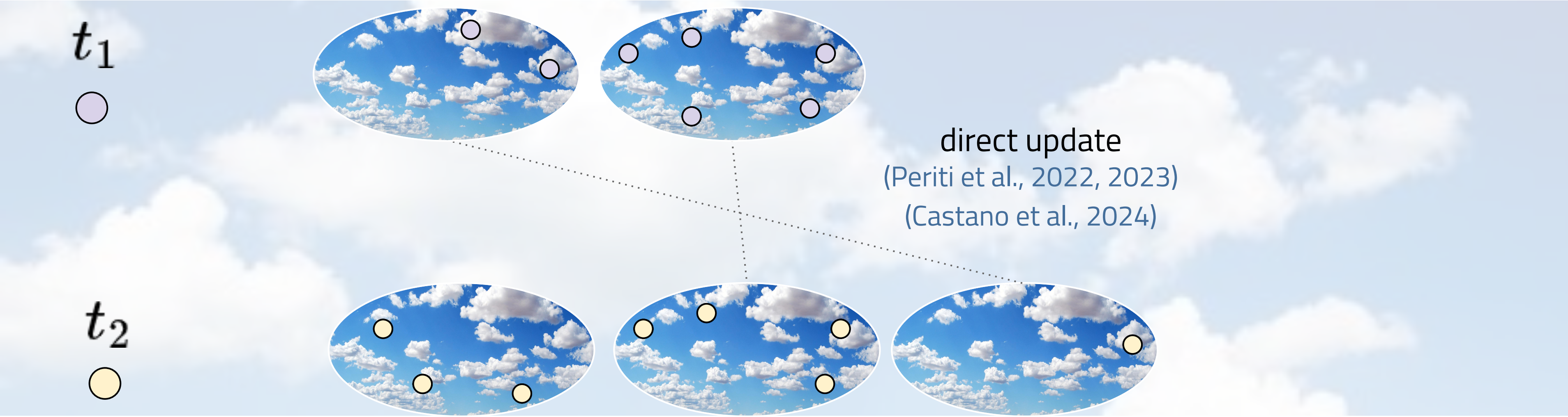
One-time clustering over all time periods

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution



Clustering over consecutive time periods

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution



Incremental/evolutionary clustering over consecutive time periods

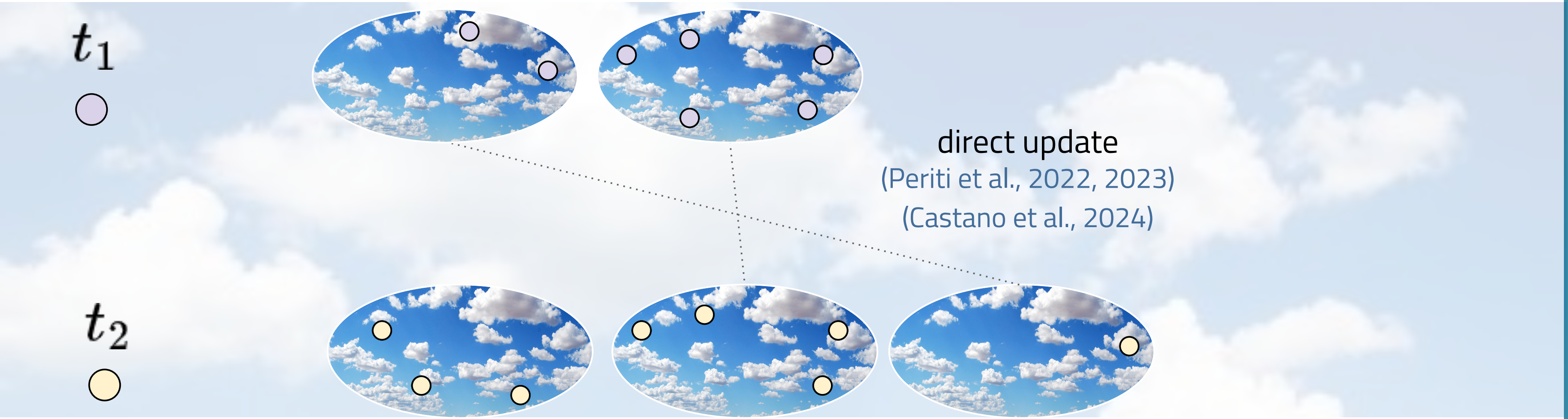
Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

- To address **interpretability** issues in a proper way
  - To better model each **individual meaning** of a word
  - To identify **type of change** by analysing the evolution of word meaning

Different approaches are suitable for different kinds of research questions and data.

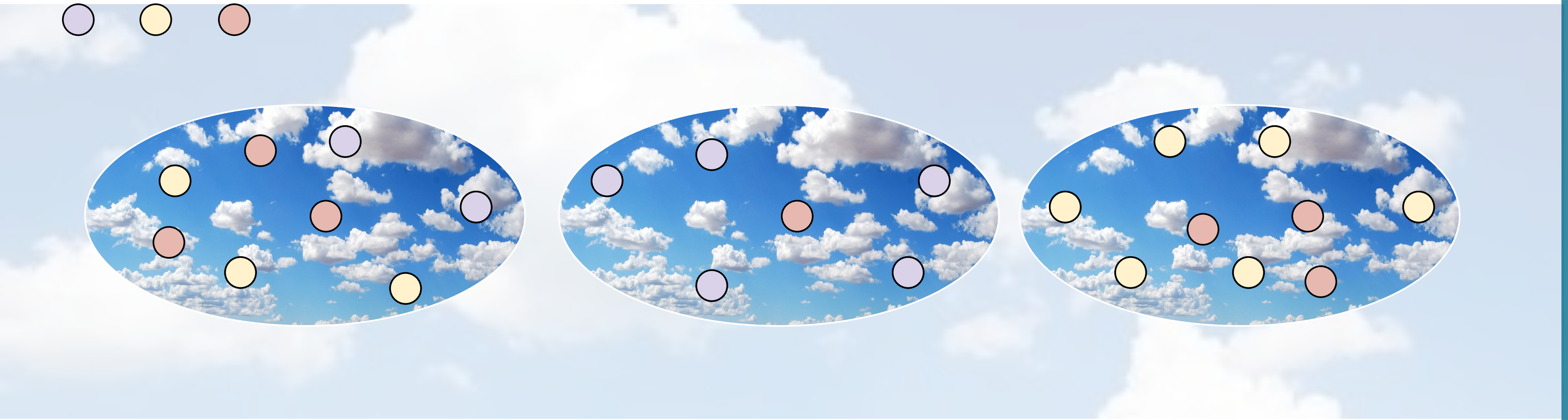
Interpretability issues arising from **contextualized LMs** word meaning interpretation and evolution



Incremental/evolutionary clustering over consecutive time periods

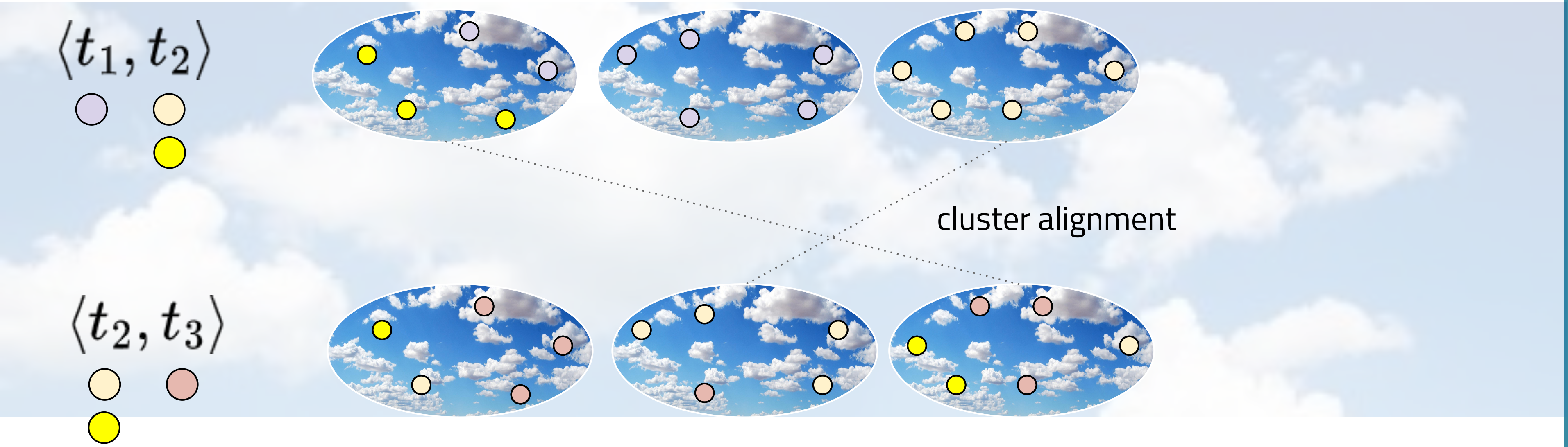
Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution

$\langle t_1, t_2, t_3 \rangle$



One-time clustering over all time periods

Interpretability issues arising from **contextualized LMs**
word meaning interpretation and evolution



Clustering is typically performed jointly (Martinc et al., 2020b)

Generalization issues arising from **contextualized LMs**
Time awareness

time-oblivious

time-aware

- contextualization

- time injection

Modern

Thou art a gay and jovial fellow.

1890

you are a cheerful and lively person.

1990

1. The context is always *time-specific.*

Google Books

Generalization issues arising from contextualized LMs
Time awareness

time-oblivious

time-aware

- contextualization
- time injection

Modern

Thou art a gay and jovial fellow.

1890

you are a cheerful and lively person.

1990

Þu eart glæd and wynsum gefera

1000

The model cannot generalize across time

Generalization issues arising from **contextualized LMs**

Time awareness

time-oblivious

- contextualization
- diachronic data

time-aware

- time injection

Thou art a gay and jovial fellow.

1890

you are a cheerful and lively person.

1990

Historical

1. The context is always *time-specific.*
2. The model is trained on *diachronic* data.

Generalization issues arising from **contextualized LMs**
Time awareness

time-oblivious

- contextualization
- diachronic data

time-aware

- time injection

Modern

Thou art a gay and jovial fellow.

1890

you are a cheerful and lively person.

1990

Þu eart glæd and wynsum gefera

1000

The model cannot generalize across time

Modeling lexical semantic change through unstructured text

(Montanelli and Periti, 2023; Tahmasebi et al., 2021; Kutuzov et al., 2018; Tang 2018)

Static Language Models

Advanced Contextualized Language Models

CIAO!

Francesco Periti

3rd year - PhD Student

Natural Language Processing, Distributional Semantic Models, Lexical Semantic Change

francesco.periti@unimi.it