



Opening up dictionaries for natural language understanding

Simon Krek (Slovenia)

Jožef Stefan Institute (Artificial Intelligence Laboratory)

University of Ljubljana (Centre for Language Resources and Technologies)

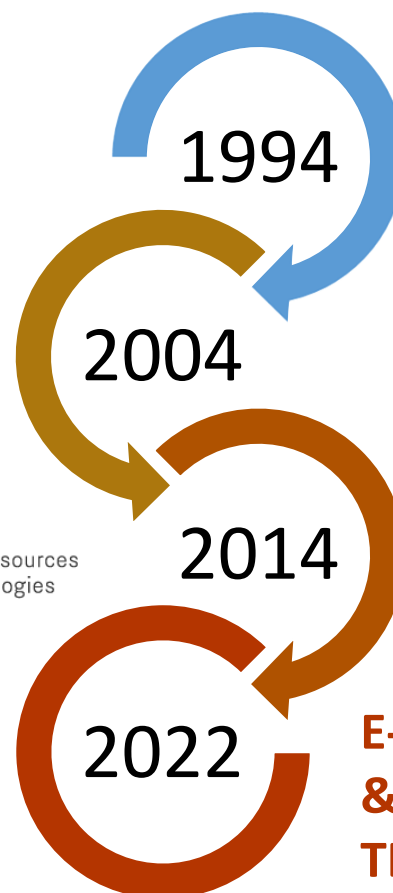
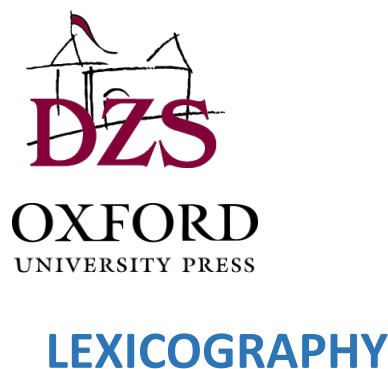


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Overview

- One slide about my interests, institutions and projects
- Benchmarking in Natural Language Understanding
- About senses and concepts (in lexicography)
- About the future of lexicography (in relation to AI / NLU)

A bit about myself



LANGUAGE
TECHNOLOGY
(FOR SLOVENE)

E-LEXICOGRAPHY
& LANGUAGE
TECHNOLOGY



Natural Language Understanding Benchmark

SuperGLUE GLUE

Paper </> Code Tasks Leaderboard FAQ Diagnostics Submit Login

MOUSSE MULTILINGUAL OPEN-TEXT UNIFIED
SYNTAX-INDEPENDENT SEMANTICS

Workshop on Ten Years of BabelNet and
Multilingual Neurosymbolic Natural
Language Understanding

4-5th JULY 2022

 **SuperGLUE**



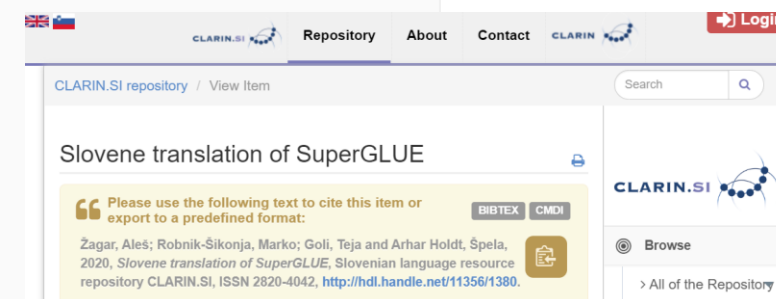
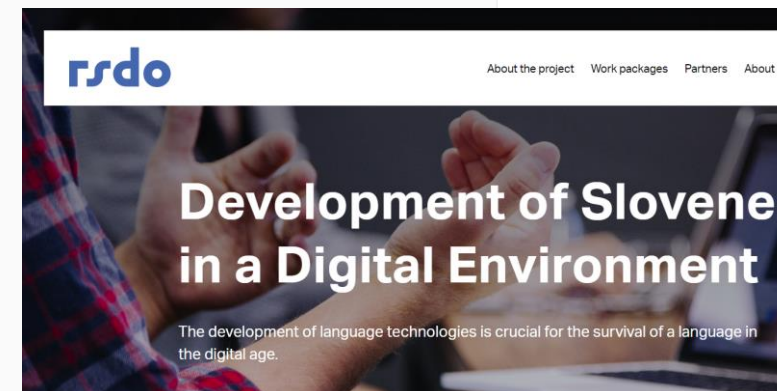
ML²























facebook Artificial Intelligence



SAMSUNG Research



SuperGLUE – 10 tasks

Name	Identifier	Download	More Info	Metric
Broadcoverage Diagnostics	AX-b			Matthew's Corr
CommitmentBank	CB			Avg. F1 / Accuracy
Choice of Plausible Alternatives	COPA			Accuracy
Multi-Sentence Reading Comprehension	MultiRC			F1a / EM
Recognizing Textual Entailment	RTE			Accuracy
Words in Context	WiC			Accuracy
The Winograd Schema Challenge	WSC			Accuracy
BoolQ	BoolQ			Accuracy
Reading Comprehension with Commonsense Reasoning	ReCoRD			F1 / Accuracy
Winogender Schema Diagnostics	AX-g			Gender Parity / Accuracy

Some examples (Wang et al. “SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems”, <https://doi.org/10.48550/arXiv.1905.00537>)

MultiRC **Paragraph:** *Susan wanted to have a birthday party. She called all of her friends. She has five friends. Her mom said that Susan can invite them all to the party. Her first friend could not go to the party because she was sick. Her second friend was going out of town. Her third friend was not so sure if her parents would let her. The fourth friend said maybe. The fifth friend could go to the party for sure. Susan was a little sad. On the day of the party, all five friends showed up. Each friend had a present for Susan. Susan was happy and sent each friend a thank you card the next week*
Question: *Did Susan’s sick friend recover?* **Candidate answers:** *Yes, she recovered (T), No (F), Yes (T), No, she didn’t recover (F), Yes, she was at Susan’s party (T)*

COPA **Premise:** *My body cast a shadow over the grass.* **Question:** *What’s the CAUSE for this?*
Alternative 1: *The sun was rising.* **Alternative 2:** *The grass was cut.*
Correct Alternative: 1

BoolQ **Passage:** *Barq’s – Barq’s is an American soft drink. Its brand of root beer is notable for having caffeine. Barq’s, created by Edward Barq and bottled since the turn of the 20th century, is owned by the Barq family but bottled by the Coca-Cola Company. It was known as Barq’s Famous Olde Tyme Root Beer until 2012.*
Question: *is barq’s root beer a pepsi product* **Answer:** No

And some more examples

ReCoRD **Paragraph:** *(CNN) Puerto Rico on Sunday overwhelmingly voted for statehood. But Congress, the only body that can approve new states, will ultimately decide whether the status of the US commonwealth changes. Ninety-seven percent of the votes in the nonbinding referendum favored statehood, an increase over the results of a 2012 referendum, official results from the State Electoral Commission show. It was the fifth such vote on statehood. "Today, we the people of Puerto Rico are sending a strong and clear message to the US Congress ... and to the world ... claiming our equal rights as American citizens, Puerto Rico Gov. Ricardo Rossello said in a news release. @highlight Puerto Rico voted Sunday in favor of US statehood*

Query For one, they can truthfully say, "Don't blame me, I didn't vote for them," when discussing the <placeholder> presidency **Correct Entities:** US

RTE **Text:** *Dana Reeve, the widow of the actor Christopher Reeve, has died of lung cancer at age 44, according to the Christopher Reeve Foundation.*

Hypothesis: *Christopher Reeve had an accident.* **Entailment:** False

WSC **Text:** *Mark told Pete many lies about himself, which Pete included in his book. He should have been more truthful.* **Coreference:** False

WiC: The Word-in-Context Dataset (English)

A reliable benchmark for the evaluation of context-sensitive word embeddings

- A system's task on the WiC dataset is to identify the intended **meaning of words**.
- WiC is framed as a binary classification task. Each instance in WiC has a target word w , either a verb or a noun, for which **two contexts** are provided. Each of these contexts triggers a specific meaning of w .
- The task is to identify if the occurrences of w in the two contexts correspond to **the same meaning** or not. In fact, the dataset can also be viewed as an application of Word Sense Disambiguation in practise.

WiC features multiple interesting characteristics:

- It is suitable for evaluating a wide range of applications, including contextualized word and sense representation and Word Sense Disambiguation;
- It is framed as a binary classification dataset, in which, unlike **Stanford Contextual Word Similarity** (SCWS), identical words are paired with each other (in different contexts); hence, a context-insensitive word embedding model would perform similarly to a random baseline;
- It is constructed using high quality annotations curated by experts.

Construction of WiC (Pilehvar & Camacho-Collados, WiC: the Word-in-Context Dataset for Evaluating Context-Sensitive Meaning Representations)

- Contextual sentences in WiC were extracted from example usages provided for words in three lexical resources:
 - (1) **WordNet** (...), the standard English lexicographic resource;
 - (2) **Verb-Net** (...), the largest domain independent verb-based resource; and
 - (3) **Wiktionary**, a large collaborative-constructed online dictionary.
- We used WordNet as our core resource, exploiting BabelNet's mappings (...) as a bridge between Wiktionary and VerbNet to WordNet.
- Lexicographer examples constitute a reliable base for the construction of the dataset, as they are curated in a way to be clearly distinguishable across different senses of a word.

WiC dataset in numbers

- Train dataset: 5,428 lines / pairs of examples
 - NOUNS: 2794 pairs / **857** headwords (in lexicographic terminology)
 - head 54
 - time 39
 - line 38
 - VERBS: 2634 pairs / **515** headwords (in lexicographic terminology)
 - make 84
 - take 55
 - break 54
- Dev dataset: 638 lines / pairs of examples
- Test dataset: 1,400 lines / pairs of examples

WiC examples (from <https://pilehvar.github.io/wic/>)

Label	Target	Context-1	Context-2
F	bed	There's a lot of trash on the <u>bed</u> of the river	I keep a glass of water next to my <u>bed</u> when I sleep
F	land	The pilot managed to <u>land</u> the airplane safely	The enemy <u>landed</u> several of our aircrafts
F	justify	<u>Justify</u> the margins	The end <u>justifies</u> the means
T	beat	We <u>beat</u> the competition	Agassi <u>beat</u> Becker in the tennis championship
T	air	<u>Air</u> pollution	Open a window and let in some <u>air</u>
T	window	The expanded <u>window</u> will give us time to catch the thieves	You have a two-hour <u>window</u> of clear weather to finish working on the lawn

Inter-annotator agreement

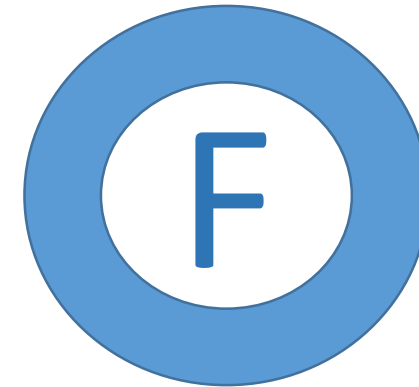
- we randomly sampled four sets of 100 instances from the test set, with an overlap of 50 instances between two of the annotators
 - each set was assigned to an annotator who was asked to label each instance based on whether they thought the two occurrences of the word referred to the same meaning or not
 - The annotators were not provided with knowledge from any external lexical resource (such as WordNet)
- We found the average human accuracy on the dataset to be **80.0%**.
 - We take this as an estimation of the human-level performance upperbound of the dataset.

WiC – the Copenhagen experiment

	Target	Context-1	Context-2
1	love (N)	Their <u>love</u> left them indifferent to their surroundings.	He hadn't had any <u>love</u> in months.
2	picture (N)	The author gives a depressing <u>picture</u> of life in Poland.	He had no clear <u>picture</u> of himself or his world.
3	line (N)	I tried to make a call, but the <u>line</u> was dead.	A dedicated <u>line</u> .
4	take (V)	Do you <u>take</u> sugar in your coffee?	A reading was <u>taken</u> of the earth 's tremors.
5	work (V)	This dough does not <u>work</u> easily.	<u>Work</u> the phones.
6	fall (V)	<u>Fall</u> asleep.	She <u>fell</u> to pieces after she lost her work.

WordNet / BabelNet ('love' as a noun)

1. bn:00031475n
a deep feeling of sexual desire and attraction
their love left them indifferent to their surroundings
2. bn:00052124n
sexual activities (often including sexual intercourse) between two people
he hadn't had any love in months
3. bn:00009819n
a beloved person: used as terms of endearment
4. bn:00052121n
a strong positive emotion of regard and affection
5. bn:00052122n
any object of warm affection or devotion
6. bn:00052123n
a score of zero in tennis or squash



Their love left them indifferent to their

He hadn't had any love in months.

‘Love’ in WiC

	Context-1	Source-1	Context-2	Source-2
T	So that's fifteen - love to Kournikova .	Wiktionary	It was 40 love .	WordNet
F	His love for his work .	WordNet	So that 's fifteen love to Kournikova .	Wiktionary
F	A mother's love is not easily shaken .	Wiktionary	The theater was her first love .	WordNet
F	Their love left them indifferent to their surroundings .	WordNet	He hadn't had any love in months .	WordNet
T	He has a very complicated love life .	WordNet	He hadn't had any love in months .	WordNet

‘Love’ in Wiktionary

1 n A deep **caring** for the existence of another.

2 n Strong affection.

2.1 A profound and caring affection towards someone.

2.2 Affectionate, benevolent concern or **care** for other people or beings, and for their well-being.

2.3 A feeling of intense **attraction** towards someone

2.4 A deep or abiding liking for something; an **enthusiasm** for something.

3 n A person who is the object of romantic feelings; a darling, a sweetheart, a beloved.

4 n (colloquial, Commonwealth of Nations) A term of friendly address, regardless of feelings.

5 n A thing, activity etc. which is the object of one's deep liking or **enthusiasm**.

6 n (euphemistic) Sexual desire; attachment based on sexual **attraction**.

7 n (euphemistic) Sexual activity.

8 n An instance or episode of being in love;

9 n Used as the closing, before the signature, of a letter,

10 n Alternative letter-case form of Love (“personification of love”).

11 n (obsolete) A thin silk material.

12 num A climbing plant, Clematis vitalba.

Etymology 2

1 n (racquet sports, billiards) Zero, no score.

6/2 ? Their love left them indifferent to their

He hadn't had any love in months.

7 ?

DANTE – A lexical database for English

1 n strong **sexual feeling** and affection for a person

2 n strong **nonsexual affection** for a person, as for one to whom one is related or with whom one has a particular friendship; also for a pet

3 n used **in** closing personal **messages** of affection

4 n [REL] the **deep commitment** and care felt by God for humankind, or for God by people, or for people because of religion

5 n strong **liking for** and **interest in** sth

6 n strong **belief in** and **commitment to** sth

7 n a **person** sb loves, a lover

8 n [spok] as a form of **address** to sb one loves or likes a lot

9 n [spok] as a familiar form of **address** to a woman whose name one does not know

10 n [inf] [BrE] a lovable or obliging **person**

11 n sth sb **likes greatly**

12 num [SPOR] [ESP TENNIS] [BADM] [esp spok] zero, no score



DANTE – A lexical database for English

1 n strong **sexual feeling** and affection for a person

- bn:00031475n
- feeling of sexual desire/attraction

2 n strong **nonsexual affection** for a person, as for one to whom one is related or with whom one has a particular friendship; also for a pet

- bn:00009819n
- beloved person

3 n used **in** closing personal **messages** of affection

- bn:00052121n
- strong emotion of regard/affection

4 n [REL] the **deep commitment** and care felt by God for humankind, or for God by people, or for people because of religion

- bn:00052122n
- object of warm affection/devotion

5 n strong **liking for** and **interest in** sth

- bn:00052124n
- sexual activities

1 Their love left them indifferent to their

He hadn't had any love in months.



DANTE – A lexical database for English

6 n strong **belief in** and **commitment to** sth

- bn:00031475n
 - feeling of sexual desire/attraction

7 n a **person** sb loves, a lover

- bn:00009819n
 - beloved person

8 n [spok] as a form of **address** to sb one loves or likes a lot

- bn:00052121n
 - strong emotion of regard/affection

9 n [spok] as a familiar form of **address** to a woman whose name one does not know

- bn:00052122n
 - object of warm affection/devotion

10 n [inf] [BrE] a lovable or obliging **person**

11 n sth sb likes greatly

- bn:00052124n
 - sexual activities



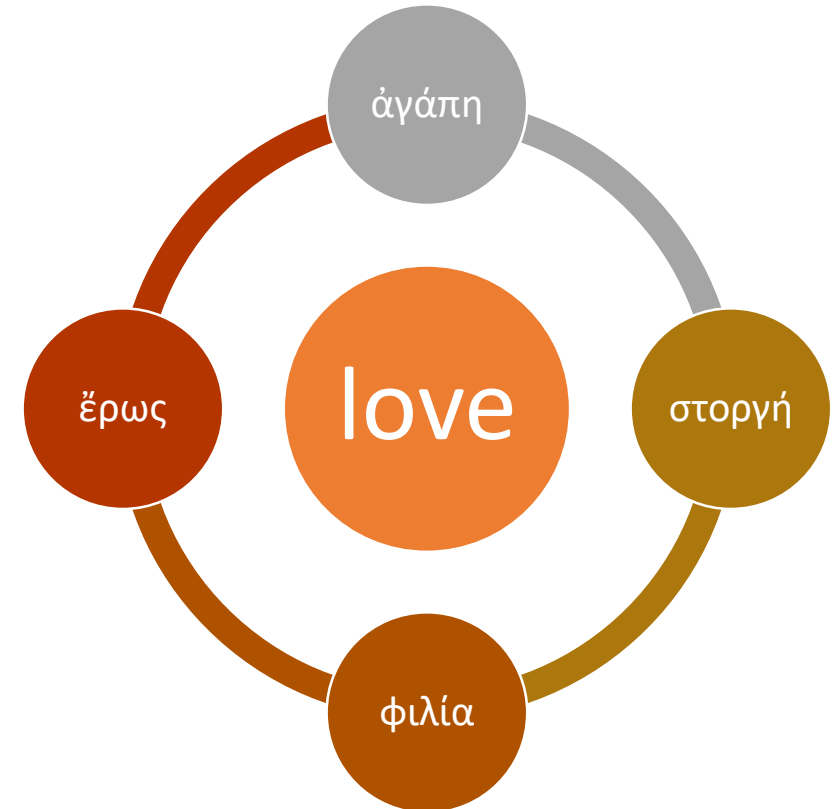
1 Their love left them indifferent to their

He hadn't had any love in months.











Let's summarize the 'love' experiment

- WordNet records six senses (synsets), the two sentences distinguish between (abstract) “sexual desire” and (concrete) “sexual activities”
- Wiktionary records twelve senses with additional four subsenses, is partly mappable to WordNet, incompatible with DANTE and quite confusing
- DANTE records twelve senses (together with phrases) but does not record “sexual activities”, senses are partly mappable to WordNet
- **Any resource (= dictionary) one takes, the task changes, more or less radically, probably with extremely low (human annotator) agreement**



SuperGLUE leaderboard

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-b	AX-g	
1	JDExplore d-team	Vega v2		91.3	90.5	98.6/99.2	99.4	88.2/62.4	94.4/93.9	96.0	77.4	98.6	-0.4	100.0/50.0	
+	2	Liam Fedus	ST-MoE-32B		91.2	92.4	96.9/98.0	99.2	89.6/65.8	95.1/94.4	93.5	77.7	96.6	72.3	96.1/94.1
3	Microsoft Alexander v-team	Turing NLR v5		90.9	92.0	95.9/97.6	98.2	88.4/63.0	96.4/95.9	94.1	77.1	97.3	67.8	93.3/95.5	
4	ERNIE Team - Baidu	ERNIE 3.0		90.6	91.0	98.6/99.2	97.4	88.6/63.2	94.7/94.2	92.6	77.4	97.3	68.6	92.7/94.7	
5	Yi Tay	PaLM 540B		90.4	91.9	94.4/96.0	99.0	88.7/63.6	94.2/93.3	94.1	77.4	95.9	72.9	95.5/90.4	
+	6	Zirui Wang	T5 + UDG, Single Model (Google Brain)		90.4	91.4	95.8/97.6	98.0	88.3/63.0	94.2/93.5	93.0	77.9	96.6	69.1	92.7/91.9
+	7	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3	90.4	95.7/97.6	98.4	88.2/63.7	94.5/94.1	93.2	77.5	95.9	66.7	93.3/93.8
8	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	76.6	99.3/99.7	
+	9	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8	88.1/63.3	94.1/93.4	92.5	76.9	93.8	65.6	92.7/91.9

What lexicographers say about dictionaries?

- Samuel Johnson (Preface, 1755):
 - 'Dictionaries are like watches. The worst is better than none, and the best cannot be expected to go quite true.'
- Manley, Jacobsen, Pedersen (Telling lies efficiently: ... 1986)
 - 'Dictionaries, monolingual and bilingual, are packs of lies: white lies, perhaps, but lies nevertheless.'
- Hanks (Lexical Analysis, 2013)
 - 'The human compulsion to make concepts tidy deserves to be studied in contrast with the essential messiness of meaning in natural language ... Such studies of a messy, mushy phenomenon should be conducted in a tidy way, of course. But the theorist's idealizations are constantly at risk of introducing gross distortion by imposing unwarranted tidiness on the data.'

Or about senses and concepts (in lexicography)?

- Atkins and Rundell: The Oxford Guide to Practical Lexicography (2008)
 - ,there is very little agreement about what word senses are or how broad their scope should be, and no definitive way of knowing when one sense ends and another begins.'
- Apresjan: Regular Polysemy (1973)
 - ,Dictionaries greatly exaggerate the measure of discreteness of meanings, and are inclined to set clear-cut borders where a closer examination... reveals only a vague intermediate area of overlapping meanings'
- Johnson (Preface, 1755)
 - ,kindred senses may be so interwoven, that the perplexity cannot be disentangled, nor any reason be assigned why one should be ranged before the other'

So, what is “humanly” possible?

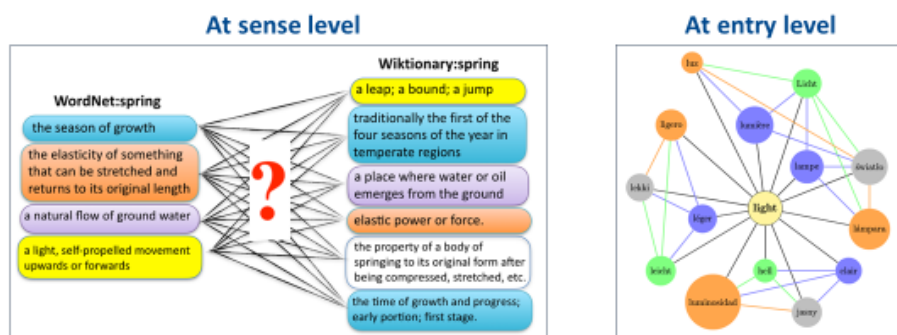
- James Murray (the first editor of OED)
 - ,the best any lexicographer could hope for would be that readers would feel, on scanning a multisense dictionary entry, that „**this is not an unreasonable way of exhibiting the facts**““ (in Atkins and Rundell)
- Dwight Bolinger (The Atomization of Meaning, 1965)
 - [Definitions] ‘help people grasp meanings, and for this purpose their main task is to supply a series of hints and associations that will relate the unknown to something known’.

ELEXIS - NAISC and BabelNet Linker

John McCrae
(ELEXIS Showcase event, Florence)

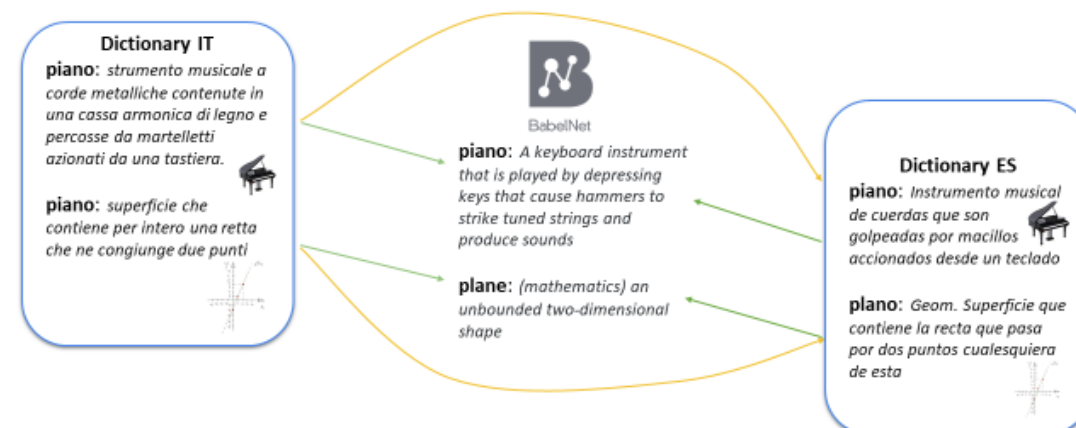
Roberto Navigli and Federico Martelli
(ELEXIS Showcase event, Florence)

Alignment at the sense and entry levels



4


Cross-lingual dictionary linking



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Lexonomy – online dictionary editor and API

Lexonomy

Dictionary 20-1page.pdf - annotate ▾ simon.krek@ijs.si ▾

Dictionaryes > 20-1page.pdf - annotate > Configure > Linking

Manual linking between entries

Elements listed here can be used as target of cross-reference link. For each element, specify unique identifier in the form of placeholders '%(element)'. Eg. element entry can have identifier '%(lemma)-%(pos)', element sense can have identifier '%(lemma)-%(number)'. Optionally, specify element you want to show as preview when selecting links.

ADD ELEMENT +

SAVE ✎

Automatic linking using NAISC

OTHER DICTIONARY CODE

START LINKING

ELEXIS community

73 institutions (17 partners, 56 observers)

ELEXIS resources

- ELEXIS collected 142 data sets
 - lexicographic resources, terminological resources, word lists, corpora, ontologies
 - [ELEXIS collection](#) in CLARIN.si repository
- 98 lexicographic resources
 - without terminology, wordnets etc.
- 50 of them with open access
- NOT ENOUGH!

Florence, June 2022



Takeaway points (1)

- Commercial lexicography is more or less gone (except perhaps for English, and in big publishing houses), remaining projects are financed by (big) technology companies or state institutions
- Much could be done by acquiring legacy knowledge (cf. ELEXIS linking) but copyright holders are extremely reluctant to make their data available, even without the real selling options (somebody else could make profit ...), as we have found in ELEXIS
 - The knowledge of (the art and craft of) lexicography, as was known in 19th/20th century, is endangered since it requires a significant personal investment to master it, but without long-term security

Takeaway points (2)

- There is a good reason why Wiktionary as a crowdsourcing project is not on the same level as Wikipedia – it will never be (also WikiData Lexicographic project). In case of polysemic lexical units, there is no clear object of description (as opposed to Named Entities), for this professional lexicographers are needed, with a well-designed manual of style.
- WordNet is widely used in NLP and AI because it is open and multilingual. In lexicographic terms, it is a problematic resource. Without an international effort to create a new open cross-lingual conceptual resource for NLU, AI will stay (semantically) blind.

Tentative prognosis (5-10 years)

- There will be an increased pressure to make (at least state-financed) lexicographic data available under open licenses (AI race)
- There will be an increased pressure to organise lexicographic data in a uniform data model (cf. OUP, also DMLex, Ontolex etc.), which will combine all types of data - morphology, lexis (also MWEs, idioms etc.), syntax, corpus data etc.
- I expect a standard for such organisation of lexicographic data will emerge bottom-up in the next 5-10 years (similar to British National Corpus in 1990s or Universal Dependencies in 2010s)
- I also expect that (open) „universal concepts and frames“ will appear



Thank you

- simon.krek@ijs.si



WiC experiment

	Target	Context-1	Context-2
F	love (N)	A mother 's <u>love</u> is not easily shaken .	The theater was her first <u>love</u> .
F	picture (N)	The author gives a depressing <u>picture</u> of life in Poland .	He had no clear <u>picture</u> of himself or his world .
T	line	I tried to make a call , but the <u>line</u> was dead .	A dedicated <u>line</u> .
F	take (V)	Do you <u>take</u> sugar in your coffee ?	A reading was <u>taken</u> of the earth 's tremors .
F	work (V)	This dough does not <u>work</u> easily .	<u>Work</u> the phones .
T	fall (V)	<u>Fall</u> asleep .	She <u>fell</u> to pieces after she lost her work .