Framing situations in the Dutch Language

A referential approach to modelling language

Piek Vossen
Computational Linguistics & Text Mining Lab (CLTL)
Faculty of Humanities
Vrije Universiteit Amsterdam

Schultink Lecture, LOT Winterschool 2022

Overview

- What is status of data for linguistics?

- How to create data for analysing linguistic framing of situations and events

Language as Data

How empirical can we get?

Language Models

- Distributional Hypothesis (Harris 1954; Firth 1957; Lenci 2008, Wittgenstein 1953): word meaning is defined by the words it co-occurs with in language use.

- “The patient is treated [MASK] [MASK]”

  [with] [penicillin, antibiotics, care, …]

  [for] [cancer, fever, herpes, mental health problems, …]

- Transformer models:

  - neural networks that learn by “statistically reading” text with “self-attention” to predict context words.

  - achieve near-human performance filling masked words in sentences or generate complete sentences

  - but also generalise over words in context such that the vector representation of “penicillin” and “antibiotics” is very similar in the sentence “The patient is treated with [MASK]”.

  - and “gazelle” gets very different representations in the next two sentences:

    - I prefer my gazelle over my giant cycling in mountains

    - The young gazelle barely escaped from the cheetah’s last jump.

BERT— Bidirectional Encoder Representations from Transformers

https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/blocks/bert-encoder

https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/

The total number of parameters is 110 million.

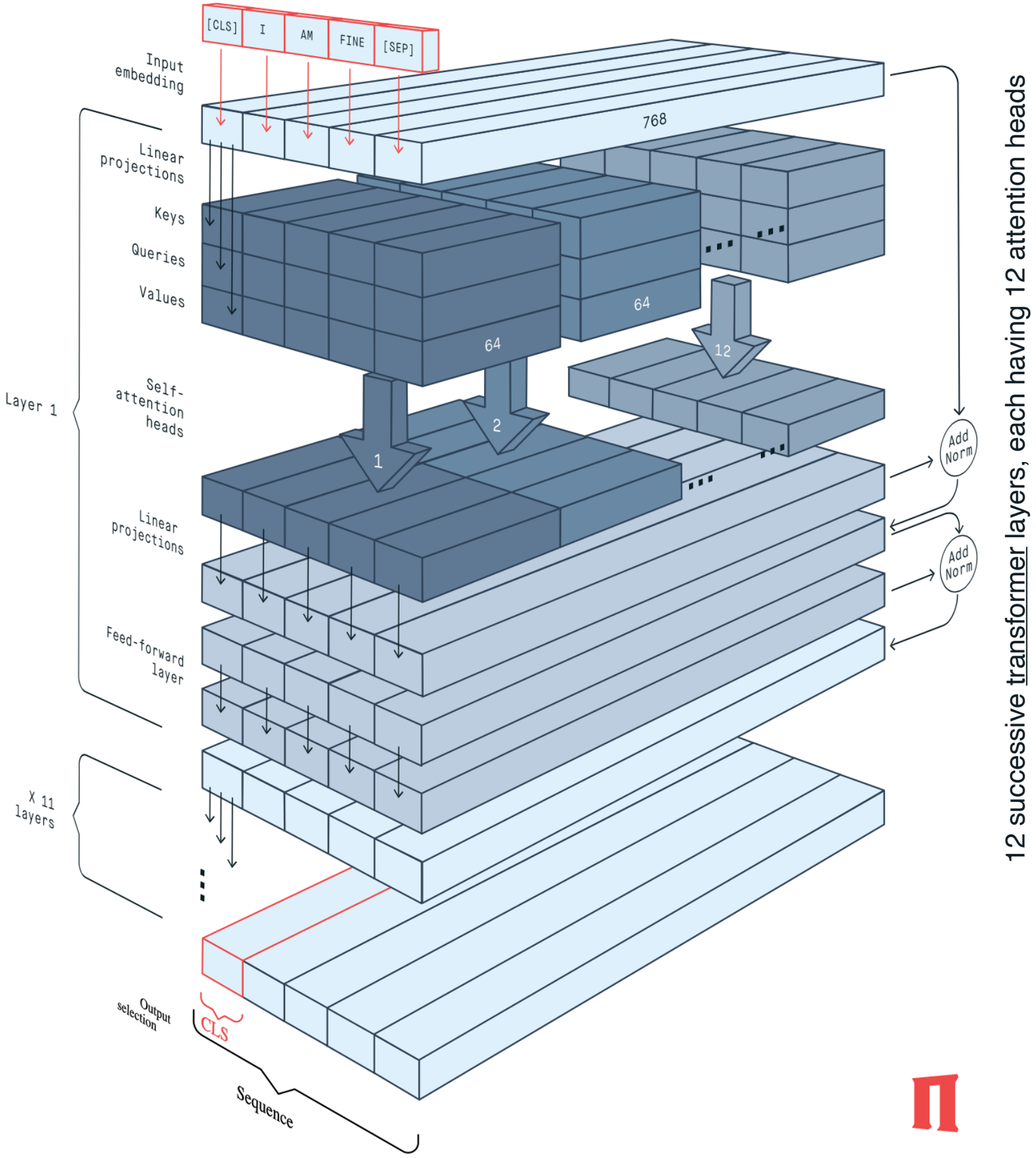
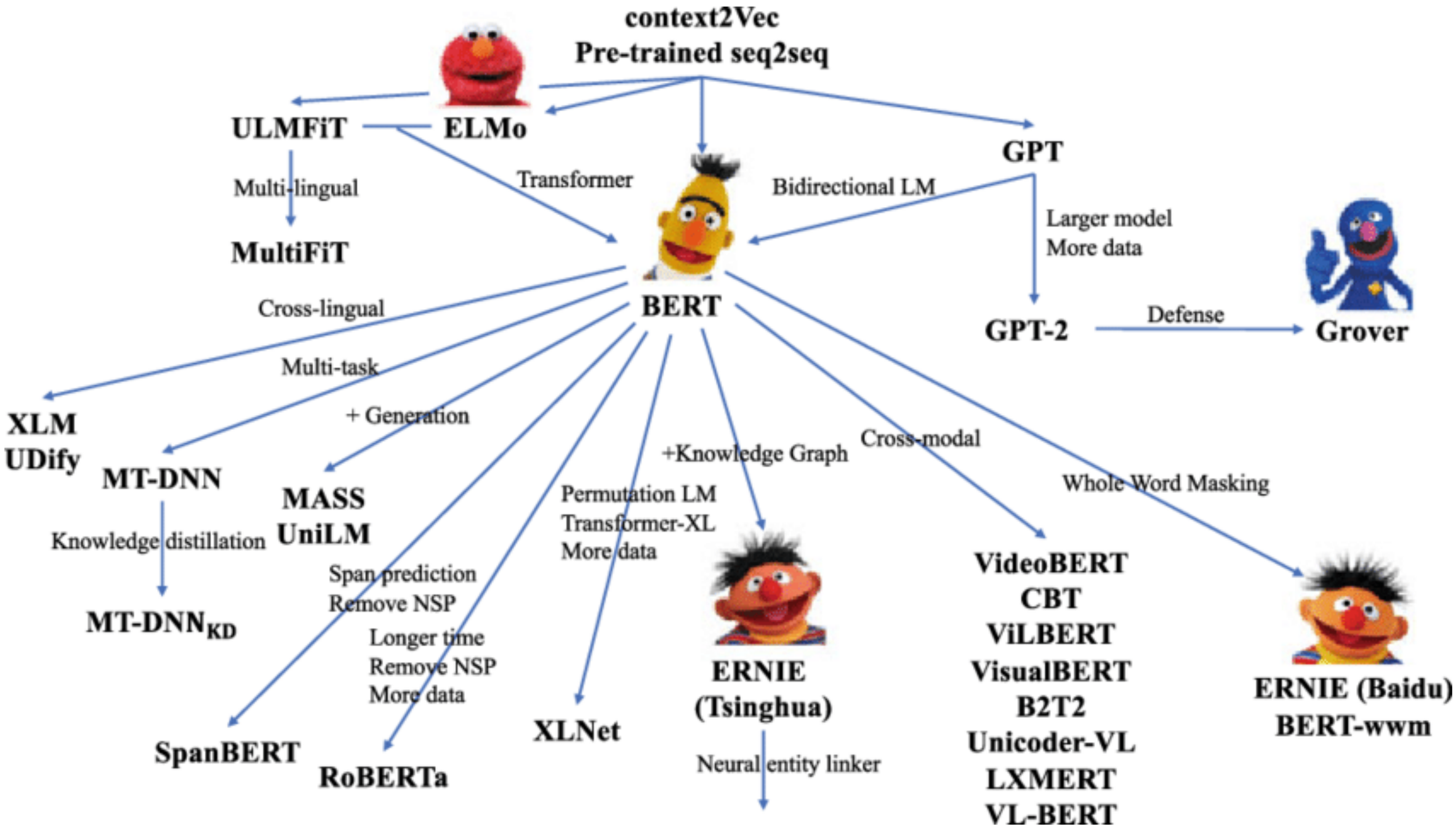


Figure 1. Structure of BERT

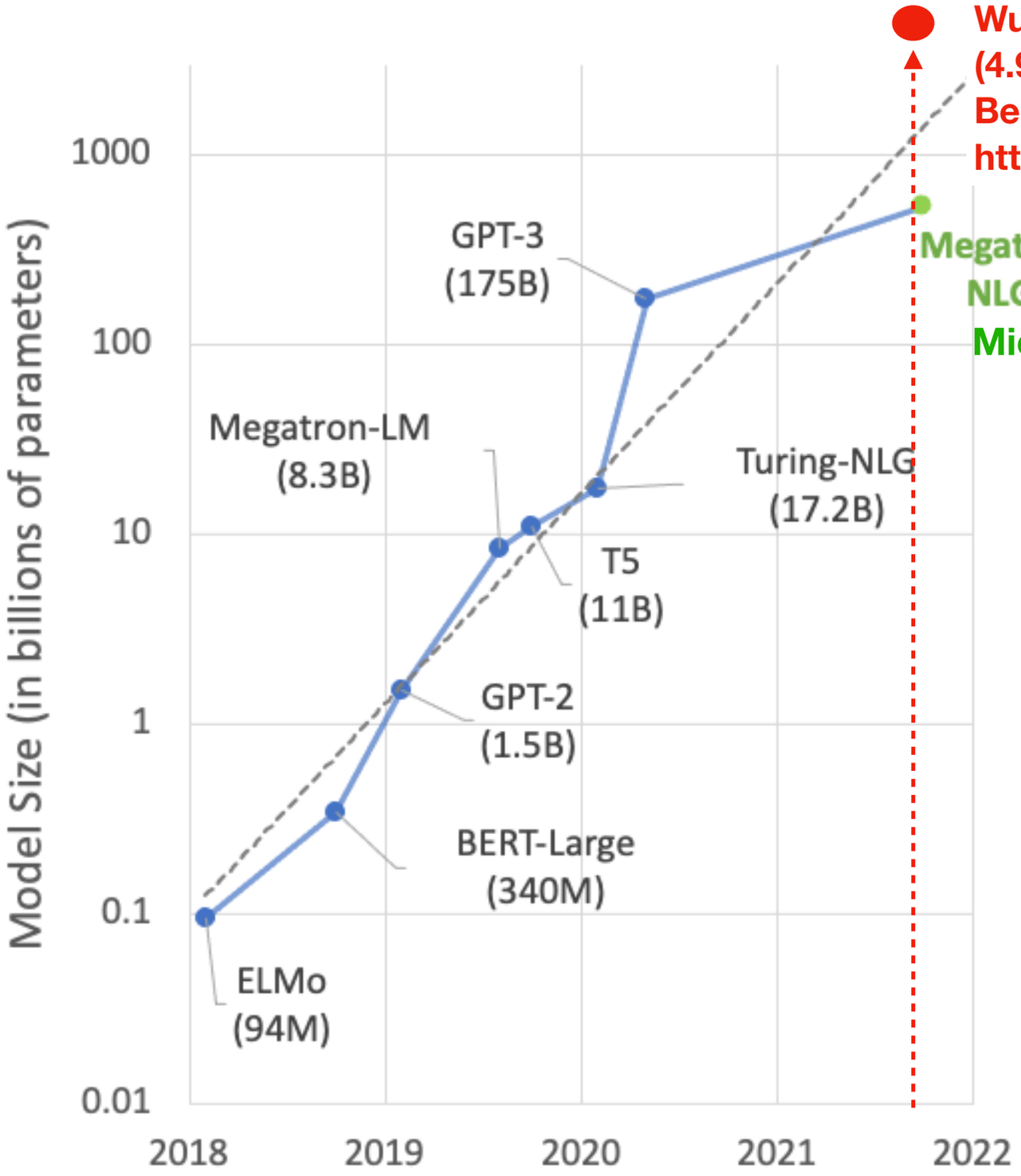
BERT's family of language models

https://huggingface.co



Liu, Zhiyuan & Lin, Yankai & Sun, Maosong. (2020). Sentence Representation. 10.1007/978-981-15-5573-2_4.

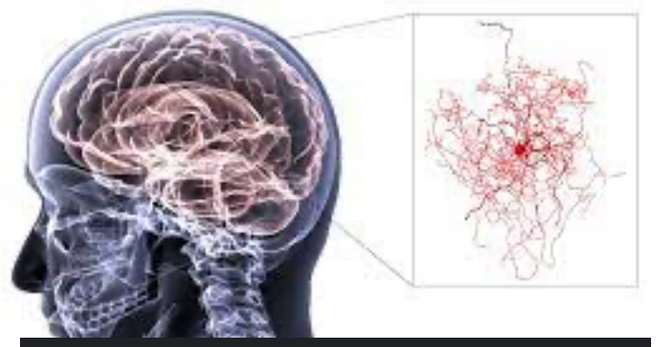
Does size matter?



Wu Dao2.0 (1.75 trillion)
(4.9 terabytes of images & texts (English/Chinese))
Beijing Academy of Artificial Intelligence
https://research.aimultiple.com/wu-dao/

Megatron-Turing
NLG (530B) 105 layers
Microsoft, NVIDIA

https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/



86 billion neurons

Dataset	Dataset source	Tokens (billions)	Weight (%)	Epochs
Books3	Pile dataset	25.7	14.3	1.5
OpenWebText2	Pile dataset	14.8	19.3	3.6
Stack Exchange	Pile dataset	11.6	5.7	1.4
PubMed Abstracts	Pile dataset	4.4	2.9	1.8
Wikipedia	Pile dataset	4.2	4.8	3.2
Gutenberg (PG-19)	Pile dataset	2.7	0.9	0.9
BookCorpus2	Pile dataset	1.5	1.0	1.8
NIH ExPorter	Pile dataset	0.3	0.2	1.8
Pile-CC	Pile dataset	49.8	9.4	0.5
ArXiv	Pile dataset	20.8	1.4	0.2
GitHub	Pile dataset	24.3	1.6	0.2
CC-2020-50	Common Crawl (CC) snapshot	68.7	13.0	0.5
CC-2021-04	Common Crawl (CC) snapshot	82.6	15.7	0.5
RealNews	RealNews	21.9	9.0	1.1
CC-Stories	Common Crawl (CC) stories	5.3	0.9	0.5

Figure 2. Datasets used to train the MT-NLG model.

Total of 339 billion tokens

Common-crawl snapshots

https://commoncrawl.org/the-data/get-started/

Turing-like test: human or machine written

GPT3 generated completions of 12 news articles (newer.com) with an average length of 215 words, where the title and subtitles is given to the system.

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**
The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Figure 3.14: The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

Brown et al, 2020: https://arxiv.org/pdf/2005.14165.pdf

Common-sense reasoning

- Semantic representations from language models are the basis for training system to succeed on common sense reasoning tasks such as the Winograde and Winogrande challenge

		Twin sentences	Options (**answer**)
✓ (1)	a	The trophy doesn't fit into the brown suitcase because **it**'s too *large*.	**trophy** / suitcase
	b	The trophy doesn't fit into the brown suitcase because **it**'s too *small*.	trophy / **suitcase**
✓ (2)	a	Ann asked Mary what time the library closes, *because* **she** had forgotten.	**Ann** / Mary
	b	Ann asked Mary what time the library closes, *but* **she** had forgotten.	Ann / **Mary**

- Sakaguchi, Keisuke, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. "Winogrande: An adversarial winograd schema challenge at scale." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8732-8740. 2020.

But there is something missing in our data

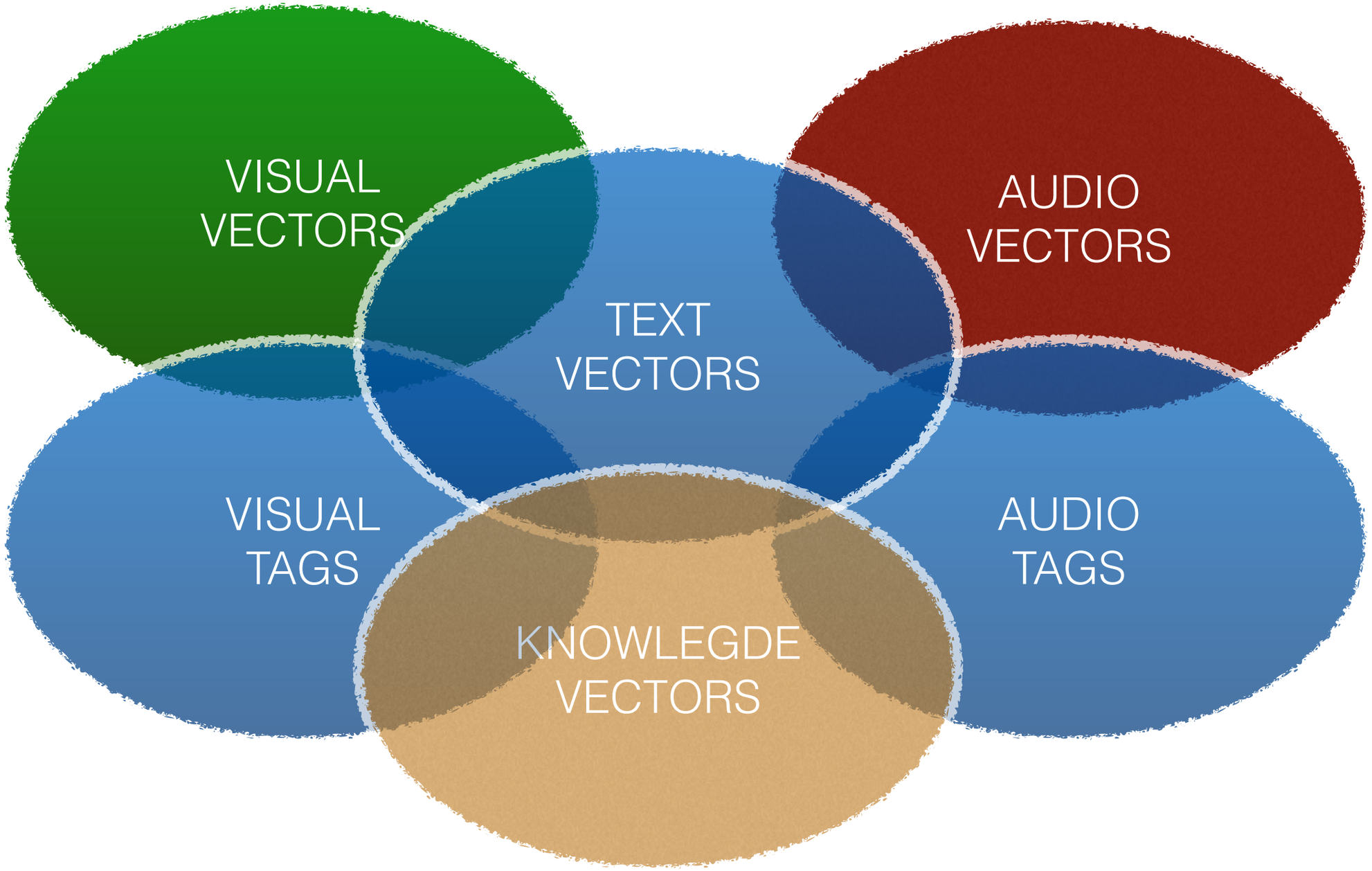
Bender, Emily M., and Alexander Koller.
"Climbing towards NLU: On meaning, form, and understanding in the age of data."
In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5185-5198. 2020.

“One approach to providing grounding is to train distributional models on corpora augmented with perceptual data, such as photos (Hossain et al., 2019) or other modalities (Kiela and Clark, 2015; Kiela et al., 2015). Another is to look to interaction data, e.g. a dialogue corpus with success annotations”, p. 5190

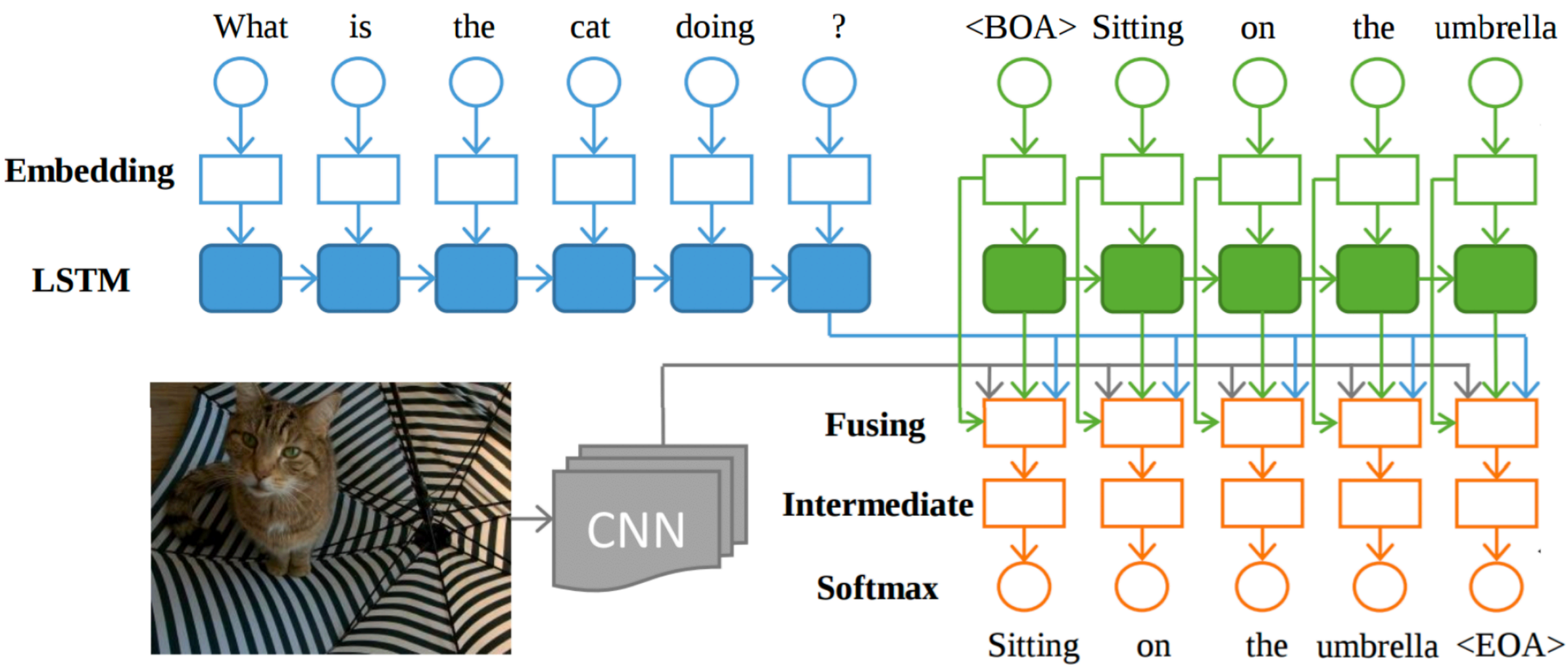
We study gazelles
by looking at gazelles
but what about
the cheetah?

We study language
by looking at language
but what about
the context of the signal?

Combining modalities



Multimodal learning



- Tangiuchi et al. 2019

Some issues

- Multimodal data is mostly static and aligned
- Lacks speech & conservation
- Grounded at the type level (*cat*) but not at the instance level (*my cat and not your cat*)
- Lacks causal/explanatory relations and narrative structure
- Lacks social, cognitive and cultural complexity

How to define context?

How to define context?

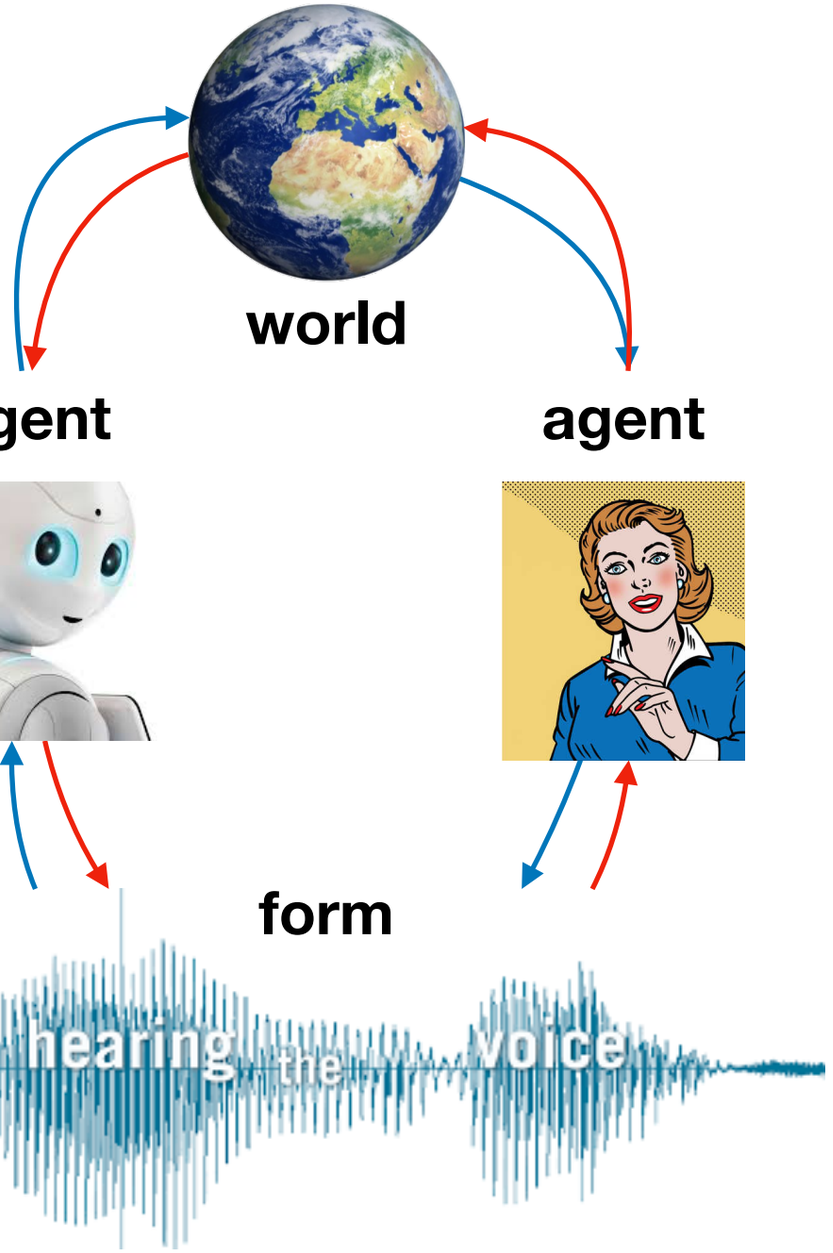
We need referential grounding

- Semantics is overrated, it is all about reference and framing (the way we make reference)

- How to explain and resolve (referential) ambiguity and variation in language in relation to context

Language Understanding & Generation By Machines

- Identity
- Reference
- Perspective



- Ambiguity
  - Sense
  - Reference
  - Vagueness
- Variation
  - Code switching
  - Pragmatics
  - Subjectivity

Worlds and Words

Sources

Signal

Concept

PLANT

PERSON

with bushy hair

Sense ambiguity

Instance

Mentions

Symbol

bush

him

president

diplomat

variation

Referential ambiguity

The world

Variation

Generation

Understanding

Ambiguity

model of the world

Referential ambiguity

- President Woodrow Wilson asked ***Ford***[?] to run as a Democrat for the United States Senate from Michigan in 1918.

- Who is Ford?
  - Gerald Ford
  - Ford, the motor company
  - Henry Ford
  - How many people in history are named “Ford”

Referential ambiguity

President [6] Woodrow Wilson [10] asked [7] Ford [8]

to run [41] as a Democrat [2] for the United States [4]

Senate [2] from Michigan [3] in 1918.

6*10*7*8*41*2*4*2*3= 6,612,480 combinations of

word senses and entities

Referential ambiguity

- President Woodrow Wilson asked ***Ford***[?] ***to run as*** a Democrat for the United States Senate from Michigan in 1918.

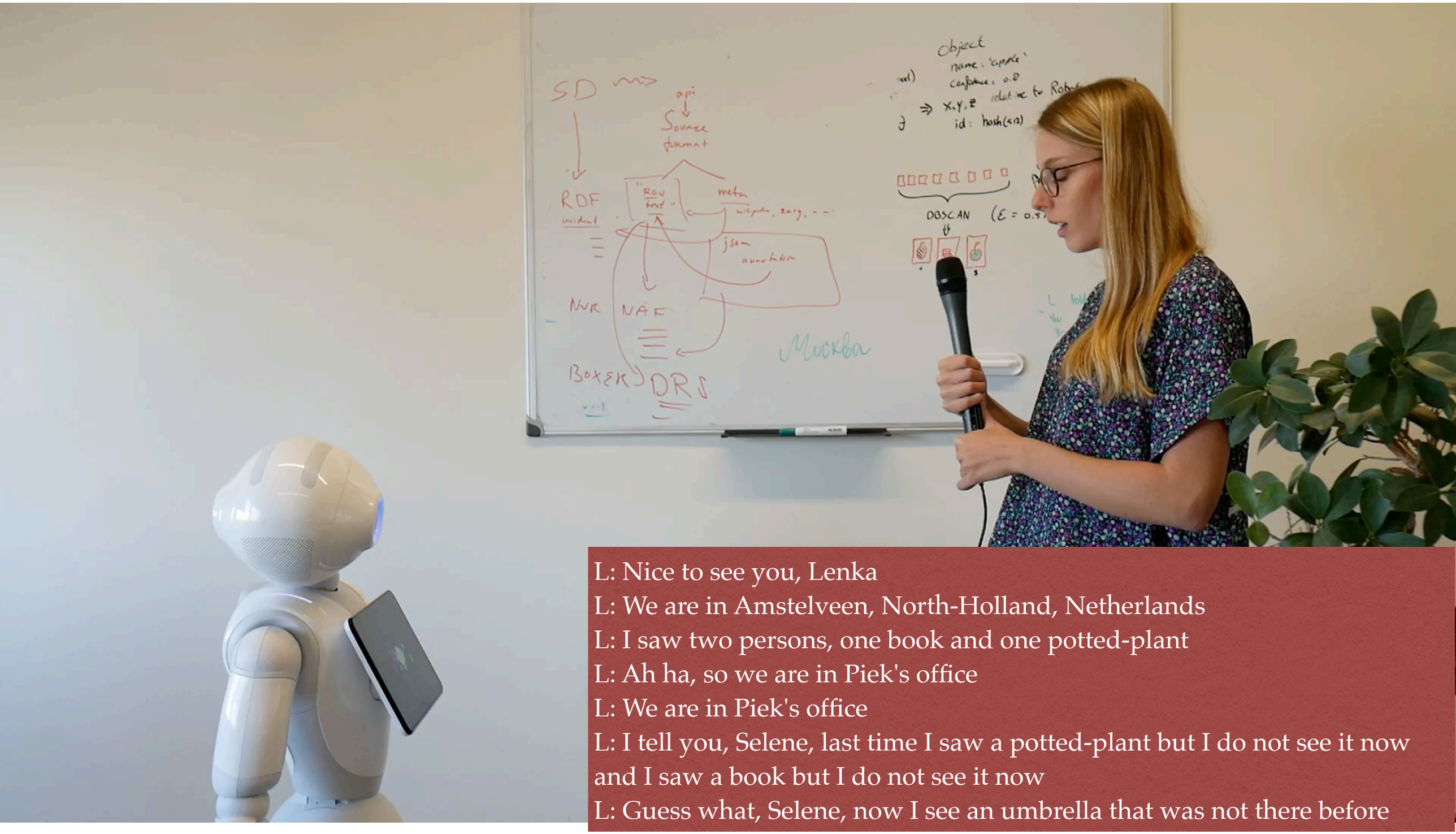
- Semantic knowledge:
  - meaning of ***to run as***

- World knowledge:
  - run as senator: +human, >18 years old, US citizen
  - Gerald Ford born in 1917, so 1 year old
  - Henry Ford the founder of Ford Motor Company, born in 1863, died in 1947, so 56 years old

Investigate ambiguity & variation empirically

- Two approaches for referential grounding:
  - Yielding referential communication in physical spaces and real time to study language interaction
    - capturing multimodal robot interaction —> makerobotstalk.nl
  - Referential grounding of stories to real world events:
    - data-to-text method in the Dutch FrameNet project —> dutchframenet.nl

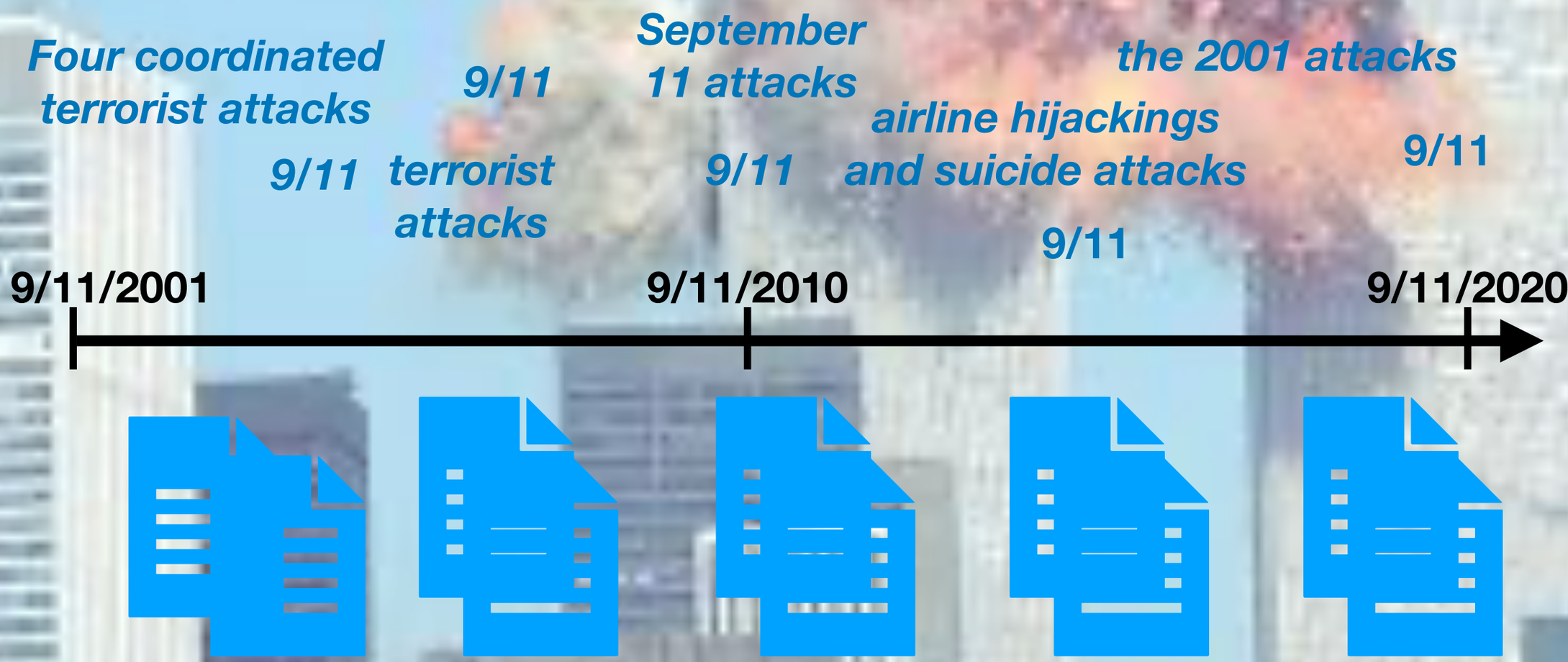
Distinguishing places and objects in contexts



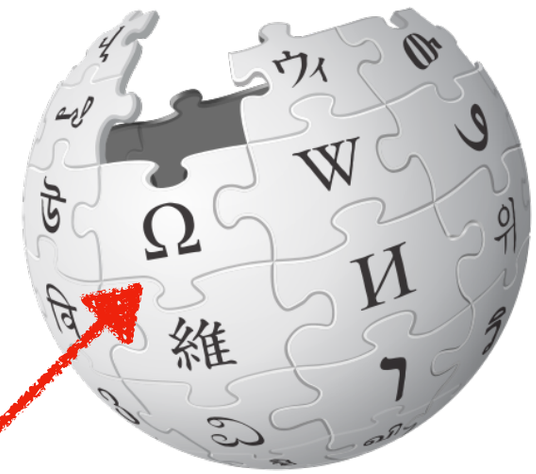
L: Nice to see you, Lenka
L: We are in Amstelveen, North-Holland, Netherlands
L: I saw two persons, one book and one potted-plant
L: Ah ha, so we are in Piek's office
L: We are in Piek's office
L: I tell you, Selene, last time I saw a potted-plant but I do not see it now and I saw a book but I do not see it now
L: Guess what, Selene, now I see an umbrella that was not there before

How to scale referentially grounded language data?

We study language
by looking at language
but what about
the context of the signal?



Referential data is very limited



- Entity and event linking & coreference
  - President Woodrow Wilson asked ***Ford*** [8] to ***run as a Democrat for the United States Senate*** [?] from Michigan in 1918. After much speculation, ***Ford*** [8] formally announced ***his*** [?] ***campaign*** [?] on June 14. Democrats were behind ***him*** [?] from the start.

Entity linking data sets

	AIDA-YAGO2	NEEL2014	NEEL2015	OKE2015	RSS500	WES2015	Wikinews
AIDA-YAGO2 (5596)	-	327 (5.87)	451 (8.06)	0 (0)	70 (1.26)	269 (4.8)	65 (1.16)
NEEL2014 (2380)	327 (13.73)	-	1630 (68.49)	57 (2.39)	61 (2.56)	294 (12.35)	67 (2.82)
NEEL2015 (2800)	451 (16.11)	1630 (58.21)	-	56 (2)	71 (2.54)	222 (7.93)	72 (2.57)
OKE2015 (531)	0 (0)	57 (10.73)	56 (10.55)	-	13 (2.44)	149 (28.06)	21 (3.95)
RSS500 (849)	70 (8.24)	61 (7.18)	71 (8.36)	13 (1.53)	-	27 (3.18)	16 (1.88)
WES2015 (7309)	269 (3.68)	294 (4.02)	222 (3.04)	149 (2.04)	27 (0.16)	-	48 (0.66)
Wikinews (279)	65 (23.30)	67 (24.01)	72 (25.81)	21 (7.53)	16 (5.73)	48 (17.20)	-

Table 7: Entity overlap in the analyzed benchmark datasets. Behind the dataset name in each row the number of uniq
dataset is given. For each datasets pair the overlap is given in number of entities and percentage (in parenthes

Ilievski, “Identity of Long-Tail Entities in Text,” PhD Thesis, 2019

Event Coreference data

Table 1: Event coreference corpora for English created by a text-to-data method

Name	Reference	nr. docs	nr mentions	mention/ docs.	nr clusters	mention/ cluster	cross doc.
ACE2005	(Peng et al., 2016)	599	5268	8.79	4046	1.30	NO
KBP2015	(Mitamura et al., 2015)	360	13113	36.43	2204	5.95	NO
OntoNotes	(Pradhan et al., 2007)	1187	3148	2.65	2983	1.06	NO
IC	(Hovy et al., 2013)	65	2665	41.00	1300	2.05	NO
EECB	(Lee et al., 2012)	482	2533	5.26	774	3.27	YES
ECB+	(Cybulska and Vossen, 2014)	982	6833	6.96	1958	3.49	YES
MEANTIME	(Minard et al., 2016)	120	2096	17.47	1717	1.22	YES
EER	(Hong et al., 2016)	79	636	8.05	75	8.48	YES
RED	(O'Gorman et al., 2016)	95	8731	91.91	2390	3.65	YES
Total		3874	36292	9.37	15057	2.41	
GVC	this publication	510	7298	14.31	1411	5.17	YES

Vossen, Ilievski, Postma and Segers, 2018

Ambiguity & Variation

MO=mean observed

A=ambiguity

V=variation

D=dominance

E=entropy

$$\frac{p(M_j,L_i)}{p(L_i)}$$

$$\frac{p(L_i,M_j)}{p(M_j)}$$

Task	Dataset	MOA	MOV	MODA	MODV	EMNLE	ELENM
EL	AIDA test B	1.09	1.35	0.98	0.91	0.05	0.22
	WES2015	1.06	1.33	0.97	0.88	0.05	0.21
	MEANTIME	1.19	4.63	0.98	0.64	0.04	0.55
EnC	QuizBowl	1.59	1.80	0.92	0.74	0.13	0.46
EvC	ECB	1.61	3.87	0.89	0.61	0.19	0.65
	ECB+	2.09	3.40	0.85	0.66	0.27	0.57
	TAC KBP '15	4.97	1.22	0.69	0.94	0.47	0.12
WSD	SE2 AW	1.20	1.06	0.94	0.98	0.13	0.05
	SE3 task 1	1.21	1.05	0.94	0.98	0.13	0.04
	SE7 task 17	1.14	1.04	0.95	0.98	0.10	0.03
	SE10 task 17	1.25	1.06	0.93	0.98	0.13	0.05
	SE13 task 12	1.10	1.06	0.97	0.98	0.14	0.05
SRL	CoNLL04	1.20	1.00	0.96	1.00	0.09	0.00

Table 11: Observed ambiguity, variance and dominance.

Entity Linking

Entity Classification

Event Coreference

Word Sense Disambiguation

Semantic Role Labeling

Ilievski, Postma and Vossen, 2016

FROM-TEXT-TO-DATA: WHICH TEXTS ARE ABOUT THE SAME EVENT?

News texts

The New York Times

United Kingdom

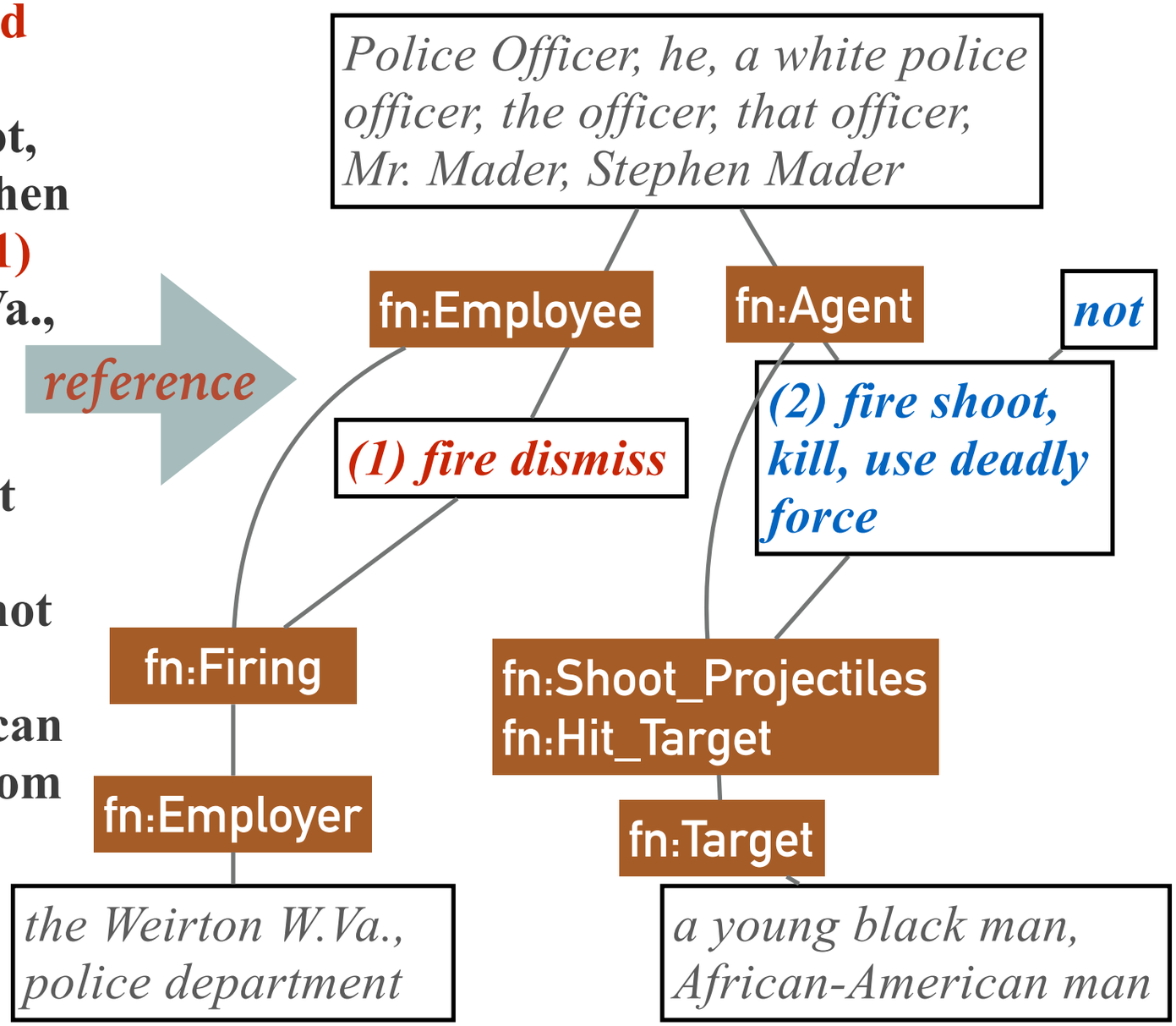
Police Officer Says He Was Fired (1) for Not Shooting (2). .. The officer could have fired (2) a shot, but he didn't. That officer, Stephen Mader, now 26, was dismissed (1) weeks later by the Weirton, W.Va., police department.

Newsweek

The Weirton Police Department terminated (1) Mr. Mader's employment because he chose not to use deadly force to shoot (2) and kill (2) and African-American man, who was suicidal, and whom Mr. Mader reasonably believed did not pose a risk of death or serious bodily injury,

reference

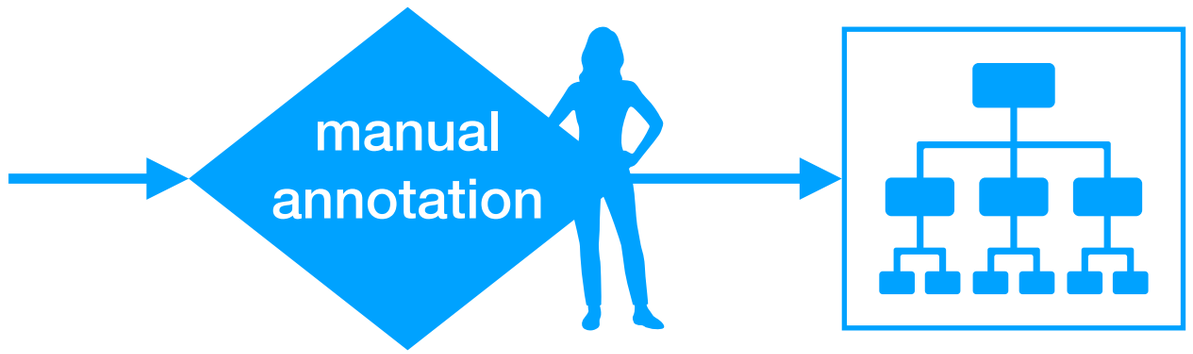
Event-Graph



Language as data

- This traditional process can be described as **text-to-data**

Police Officer Says He Was Fired (1) for Not Shooting (2). .. The officer could have fired (2) a shot, but he didn't. That officer, Stephen Mader, now 26, was dismissed (1) weeks later by the Weirton, W.Va., police department.



- We propose a new method: **data-to-text**, which helps us to get from text-to-data
- We apply this method to learn the language of framing situations across different languages

1. Sparql query for accidents

Wikidata Query | Examples | Help | Tools

Query Helper

instance of | accident

any

subclass of

+ Filter

+ Show

Limit 90

```
#SELECT (count(distinct ?event) as ?cnt)
SELECT ?event ?eventLabel
WHERE
{
    # find events
    ?event wdt:P31/wdt:P279* wd:Q171558.
    #?event wdt:P31 wd:Q83267.
    # with a point in time or start date
    OPTIONAL { ?event wdt:P585 ?date. }
    OPTIONAL { ?event wdt:P580 ?date. }
    # but at least one of those
    FILTER(BOUND(?date) && DATATYPE(?date) = xsd:dateTime).
    # not in the future, and not more than 31 days ago
    BIND(NOW() - ?date AS ?distance).
    FILTER(0 <= ?distance && ?distance < 36500).
    # and get a label as well
    OPTIONAL {
        ?event rdfs:label ?eventLabel.
        FILTER(LANG(?eventLabel) = "en").
    }
}
# limit to 10 results so we don't timeout
LIMIT 90
```

2. List of accidents

wd:Q223288	Tenerife airport disaster
wd:Q265922	KLM Flight 867
wd:Q275701	2012 Trans Air Congo Ilyushin Il-76 crash
wd:Q282999	Aero Caribbean Flight 883
wd:Q308923	Mumbai
wd:Q314087	
wd:Q320879	
wd:Q335259	
wd:Q337044	
wd:Q346847	
wd:Q348349	

3. Structured data for accident

WIKIDATA

Item | Discussion | Read | View history

Tenerife airport disaster (Q223288)

deadliest aviation accident ever

Tenerife Disaster

edit

In more languages Configure

Language	Label	Description	Also known as
English	Tenerife airport disaster	deadliest aviation accident ever	Tenerife Disaster
Dutch	Vliegtuigramp van Tenerife	No description defined	Vliegtuigramp Tenerife Teneriferamp
German	Flugzeugkatastrophe von Teneriffa	Flugzeugkollision im Jahr 1977	Flugzeugunglück von Teneriffa KLM-Flug 4805 PanAm-Flug 1736
French	catastrophe de Ténérife	accident d'avions le plus meurtrier de l'Histoire	Désastre de Tenerife Catastrophe de Tenerife Désastre de Ténérife Désastre de Ténériffe

All entered languages

Statements

instance of — aviation accident — edit

0 references

add reference

add value

image — KLM 747 (7491686916).jpg — edit

1 reference

add value

Main page
Community portal
Project chat
Create a new item
Recent changes
Random item
Query Service
Nearby
Help
Donate

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

Wikipedia (38 entries) edit

als Flugkatastrophe von Teneriffa
ar كارثة مطار تنريف
az Tenerife aeroportunda qəlaqət (2
bg Авиокатастрофа на Тенерифе 1977 г.
bn তেনেরিফে বিমানবন্দর বিপর্যয়
ca Desastre de Tenerife
cs Letecké neštěstí na Tenerife 197
da Flyulykken på Tenerife
de Flugzeugkatastrophe von Teneri
en Tenerife airport disaster
eo Aviadila kolizio de 1977 en Los
es Accidente de Los Rodeos
et Tenerife lennukatastroof
fa حادثه فرودگاه تنریف
fi Teneriffan lento-onnettomuus
fr Accident aérien de Tenerife
gl Accidente de Los Rodeos
he אסון התעופה בטנריף
hr Zrakoplovna nesreća na Tenerifi
hu Tenerifei katasztrófa
id Musibah Tenerife
it Disastro aereo di Tenerife
ja テネリフェ空港ジャンボ機衝突
ko 테네리페 참사
nl Vliegtuigramp van Tenerife

4. Link to Dutch Wikipedia

5. Dutch Wikipedia

WIKIPEDIA
De vrije encyclopedie

Hoofdpagina
Vind een artikel
Vandaag
Etalage
Categorieën
Recente wijzigingen
Nieuwe artikelen
Willekeurige pagina

Informatie
Gebruikersportaal
Snelcursus
Hulp en contact
Donaties

Hulpmiddelen
Links naar deze pagina
Verwante wijzigingen
Bestand uploaden
Speciale pagina's
Permanente koppeling
Paginagegevens
Wikidata-item
Deze pagina citeren

Afdrukken/exporteren
Boek maken
Downloaden als PDF
Printvriendelijke versie

In andere projecten
Wikimedia Commons

In andere talen
العربية
Deutsch
English

Artikel | Overleg

Vliegtuigramp van Tenerife

De **vliegtuigramp van Tenerife** vond plaats op zondag 27 maart 1977, toen op de luchthaven Tenerife Noord (destijds bekend a Tenerife twee vliegtuigen van het type Boeing 747 – een van Pan American World Airways (*Pan Am*) en een van KLM – op elkaa het leven. Het is daarmee de grootste ramp uit de geschiedenis van de luchtvaart. Slechts 61 inzittenden, allen aan boord van he

Inhoud [verbergen]

1 Omstandigheden
2 Fatale gebeurtenissen
  2.1 Afslag gemist
  2.2 Dichte mist
  2.3 Overhaaste start, misverstanden en radiostoring
  2.4 Mislukte uitwijkmanoeuvres, botsing, en brand
3 Onderzoek en maatregelen
4 Monumenten en herdenking
5 Varia
6 Externe links

Omstandigheden [bewerken]

Op 27 maart 1977 was de Pan Am-vlucht 1736 anderhalf uur te laat uit Los Angeles vertrokken naar de Canarische Eilanden, m eens 103 passagiers instapten. Aan boord waren in totaal 378 Amerikaanse passagiers voor een cruise op de Middellandse Zee eindbestemming Las Palmas de Gran Canaria naderde, kreeg gezagvoerder Victor Grubbs te horen dat de luchthaven tijdelijk w door een door Antonio Cubillo geleide afscheidingsbeweging en telefonische dreiging met een tweede aanslag. Grubbs wilde bli open ging. Dit werd echter geweigerd en de vlucht werd omgeleid naar de luchthaven *Los Rodeos* op het 110 kilometer noordwe de hoofdstad Santa Cruz de Tenerife. Daar kwam het toestel aan rond 14.00 uur plaatselijke tijd. Het moest wachten tot er verde Tenerife gingen twee personeelsleden van Pan Am als passagier aan boord; zij namen plaats in de cockpit.

KLM-vlucht 4805, een chartervlucht vol met vakantiegangers van Holland International, voornamelijk Nederlanders, was dezelfde Las Palmas. Ook dit toestel, de KLM-Boeing *De Rijn*, moest uitwijken naar Tenerife. Het was daar circa 25 minuten vóór de Pan

Rond 15.00 uur werd Las Palmas weer vrijgegeven. Er waren geen andere bommen gevonden. Het probleem was nu dat een de met toestellen die waren uitgeweken van Las Palmas. Daarom moest de startbaan ook als taxibaan gebruikt worden.

Het KLM-toestel stond om 16.50 uur op het punt om door te reizen naar Las Palmas. De gezagvoerder had het toestel laten volta dens het tanken blokkeerde het KLM-toestel de doorgang van h de Pan Am-Boeing zou direct daarna vertrekken. De instructies

6. Links to reference texts

Externe links [bewerken]

- (es) Webversie van het officiële Spaanse rapport
- (en) Engelse vertaling van het Spaanse rapport, inclusief commentaar hierop van
- Definitief rapport van de Raad voor de Luchtvaart:
  - Nederlands
  - Engels
- (en) Definitief rapport van de Air Line Pilots Association
- International Tenerife Memorial
- Project Tenerife

Bronnen

1. ↑ International Tenerife Memorial, tenerife-memorial.org
2. ↑ Afbeelding op project-tenerife.com
3. ↑ (en) Air Crash Investigation The Tenerife airport disaster op YouTube
4. ↑ Volkskrant.nl Geboren voor het ongeluk

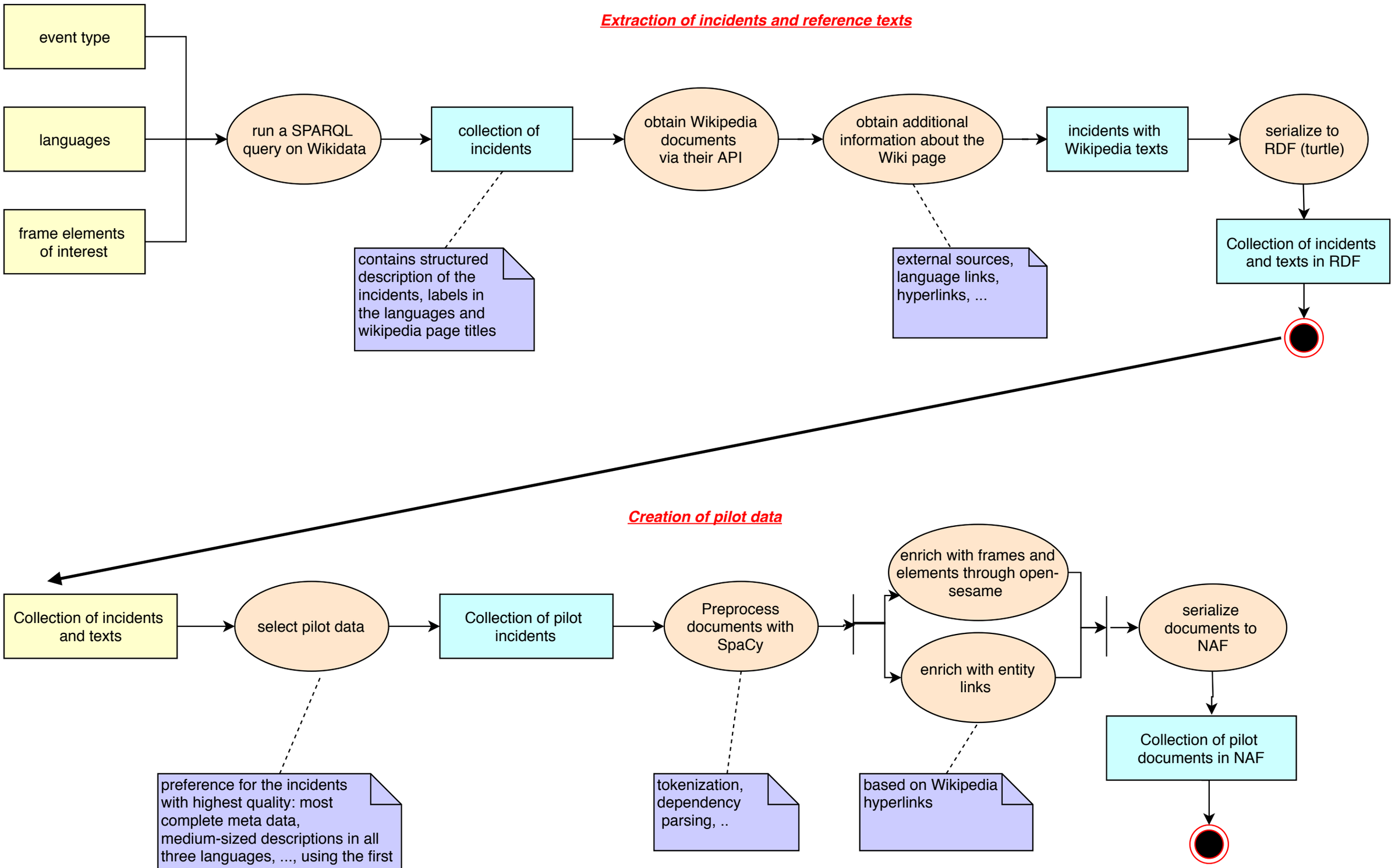
7. Reference text original

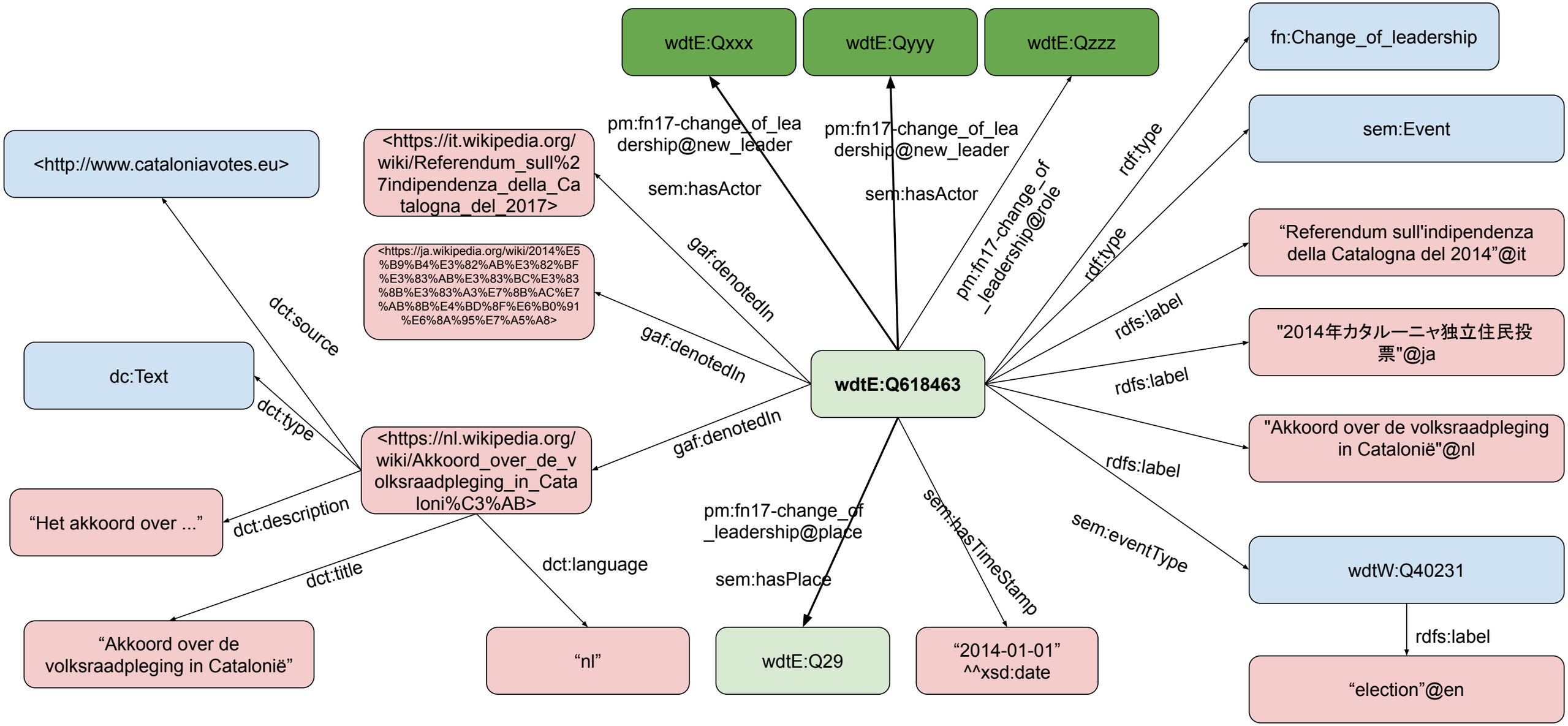
https://www.volkskrant.nl/archief/geboren-voor-het-ongeluk~a3307696/

8. Reference text in the wayback machine

https://web.archive.org/web/20171012150057/https://www.volkskrant.nl/archief/geboren-voor-het-ongeluk~a3307696/

Extraction Process

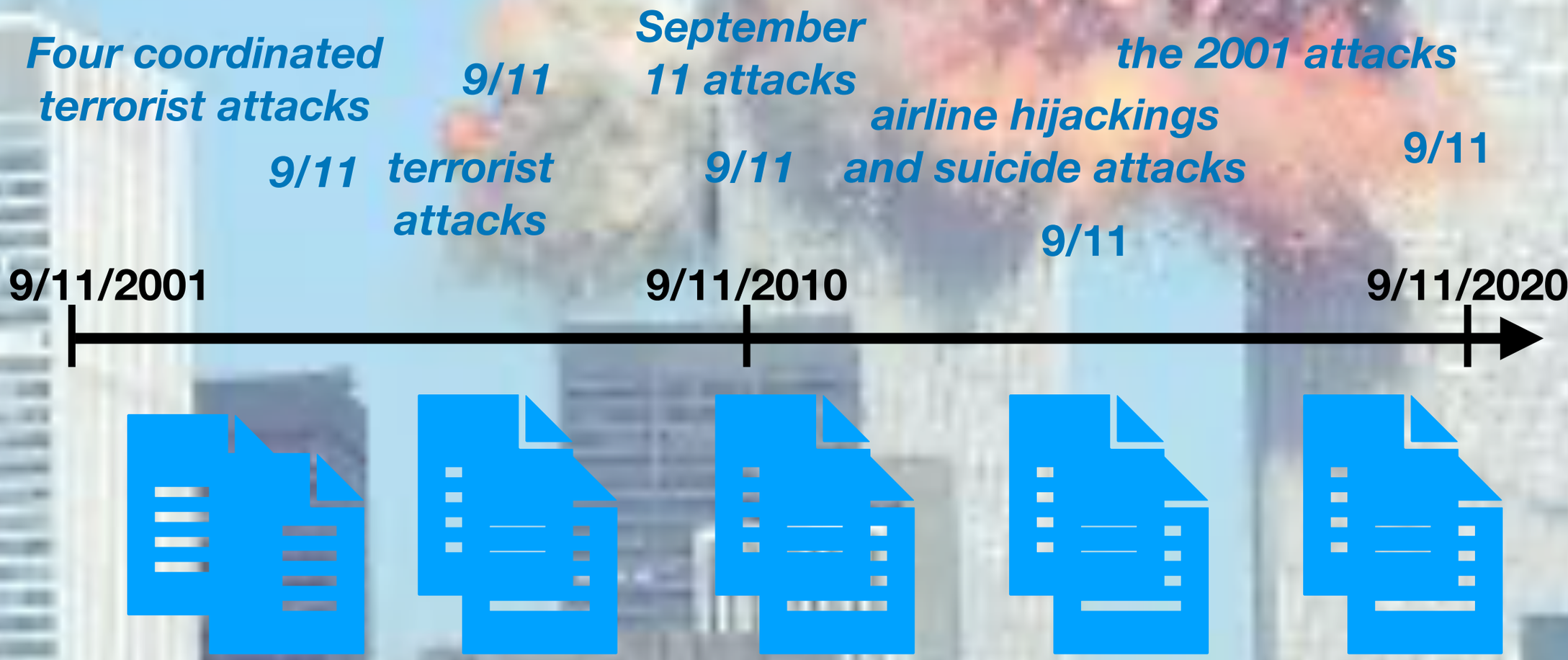




Dublin Core (**dc**): *meta data on sources*
FrameNet (**fn**): *conceptual situational schema*
Simple Event Model (**sem**): *events, participants, time and location relations*
Grounded Annotation FrameWork (**gaf**): *anchoring instances to mentions in sources*
Others (**rdfs, rdf, owl time**)

Now we have texts that are referentially grounded (no ambiguity),

how to understand referential variation?



A text tells a story

A murder and conviction: *what, who , when and where*

- Jury **convicts** man in woman's death Saturday , 28 September 2013 01 : 43. A jury in eastern Oklahoma has **convicted** a Spiro man of two counts of first - degree **murder** in the 2012 **shooting death** of his **pregnant** girlfriend . The jury **deliberated** almost seven hours Thursday before **convicting** 27 - year - old Christopher Kenyon Simpson in the death of 20 - year - old Ka'loni Flynn , of Fort Smith , Ark . The jury **recommended** the maximum **sentence** of life in prison without parole.

FrameNet

- Variation reflects framing
- Words ***evoke*** a situational Frame with encyclopaedic knowledge
- Situation dependent roles Perpetrator, Crime, Victim, Offense, Verdict, Sentence
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. "The berkeley framenet project." 1998.

Charles Filmore

Do case roles define the predicate or does the predicate define the case roles?

A text tells a story
what, who, where, when

Judge | Verdict | Defendant | Charges — **FrameNet situation**

Agent | Predicate | Patient — **Semantic roles**

Subject | Verb | Object | Adjunct — **Syntactic function**

- **Jury convicts man in woman's death** Saturday , 28 September 2013 01 : 43.

FrameNet database

- **Frame**: **Offense**
- **Frame Elements**: **Perpetrator, Victim**
- **Lexical Units**: **arson.n, assault.n, battery.n, burglary.n, child abuse.n, conspiracy.n, copyright infringement.n, felony.n, fraud.n, hijacking.n, homicide.n, indecent assault.n, kidnapping.n, larceny.n, manslaughter.n, murder.n, negligence.n, possession.n, rape.n, robbery.n, sabotage.n, sexual assault.n, sexual harassment.n, statutory rape.n, theft.n, treason.n**

- **Frame**: **Arson** — inherits-from —> **Offense**
- **Lexical Units**: **arson.n, arsonist.n**

FrameNet Status

- > 20 years of building
- 1,087 frames
- 10,542 frame elements
- 13,640 lexical units in English
- 202,232 annotations in English texts
- FrameNet in other languages: French, Swedish, Japanese, Portuguese, Chinese, German, Spanish, Korean
- Global fragment project: https://www.globalframenet.org

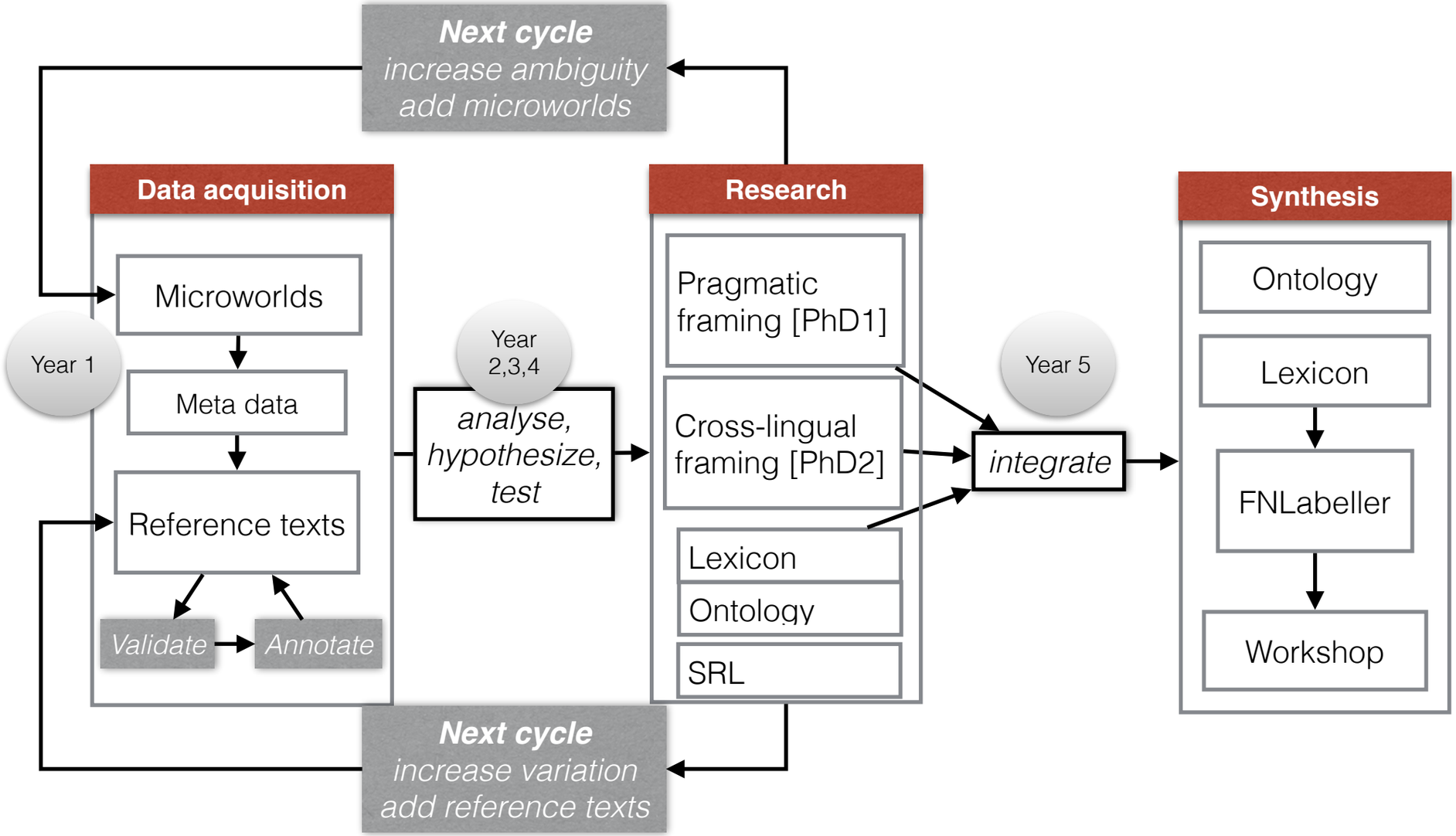
Framing situations in the Dutch Language

from structured data to text
and back from text to structured data on situations

Piek Vossen, Marten Postma, Filip Ilievski, Antske Fokkens, VU University Amsterdam
Johan Bos, Malvina Nissim, Tommaso Caselli, Groningen University

http://dutchframenet.nl

Dutch FN Overview



Project phases

1. Create large data sets with microwords and reference texts for various event types
2. Pilot project to discover typical frames per event type
3. First annotation round
    1. Pre-annotation: lexical and fine-tuned transformer
    2. *Analysing variation and pragmatic factors*
4. *Second annotation round*
    1. *Pre-annotation: lexical and fine-tuned transformer*
    2. *Analysing variation and exploring pragmatic factors*
5. *Third annotation round*
    1. *Automatic annotation: fine-tuned transformer*
6. *Final lexicon, data and corpora*

Annotation

- Structured data and (cross-lingual) texts referentially grounded to incidents for a specific type of events —> consistency across texts

- Typical frames for event type —> consistency and relevance across incidents

- Annotation decisions:

1. Link mentions to structured data of an incident (entities and events)

2. Assign frames and frame elements for linked events —> main narrative

3. Assign frames to mentions of subevents (included in the temporal & causal container of the main event, O'Gorman et al., 2016; Caselli and Vossen, 2017)

4. Core elements outside sentence —> first mention or mark as unexpressed

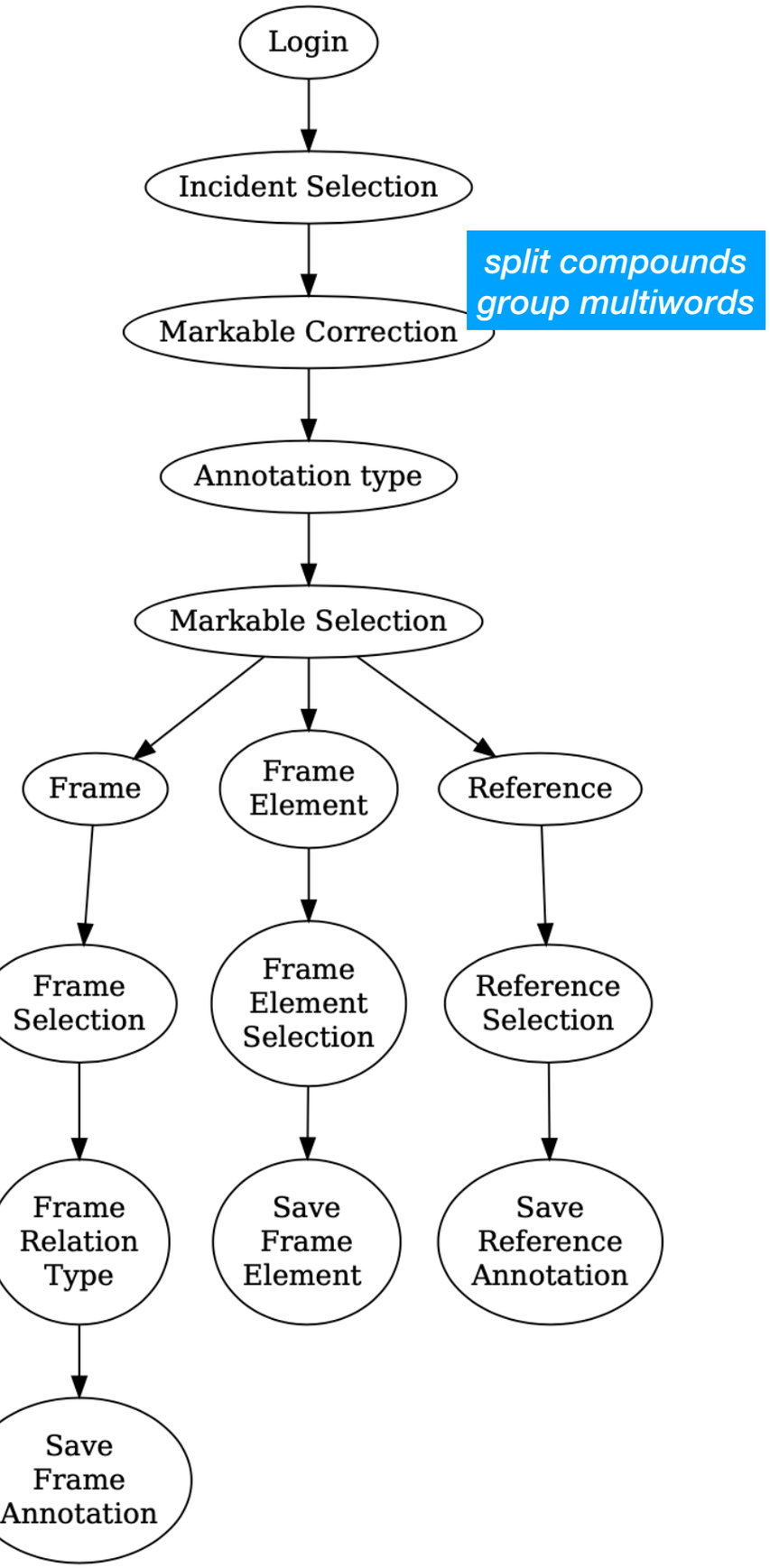


Figure 4: Annotation workflow

Frame Annotation Tool

http://frame.cltl.nl/annotation

Logged user: **piek**

Back to dashboard

Logout

test_release | aircraft shootdown@ | Malaysia Airlines Flig | nl | 298 doden bij Malays

Select Annotation Type: Frame Annotation

Select Annotation Task: Remove

Select Typicality range:

Save | Clear selection

298 doden bij Malaysia-crash (source)

In Oekraïne is een passagier_s_**vliegtuig** van Malaysia Airlines **neergestort** dat in Amsterdam is **opgestegen** . Malaysia Airlines bevestigt dat **vlucht** MH17 wordt **vermist** . Dat **gebeurde** in de buurt van Donetsk op 50 kilometer van de Russische grens . Malaysia Airlines zegt in een verklaring dat het toestel 283 passagiers en 15 bemanningsleden **aan boord had** . Daar **zitten** veel Nederlanders **bij** , zei minister Opstelten . Hoeveel is nog onduidelijk . Het toestel was om 12.14 uur van Schiphol **opgestegen** . Het zou om 00.10 Nederlandse tijd in Kuala Lumpur **aankomen** . KLM Vlucht MH17 van Malaysia Airlines is een gecombineerde lijndienst van Malaysia Airlines en KLM , die dagelijks op Kuala Lumpur vliegt . Het KLM - vluchtnummer was KL4103 . KLM zegt in een verklaring dat het met leedwezen kennis heeft genomen van " een mogelijk **incident** " met deze **vlucht** . " We staan in contact met Malaysia Airlines om meer informatie te krijgen . " Schiphol heeft een GRIP2-procedure in werking gezet waardoor er hulp_**verleners** naar de luchthaven zijn **gekomen** . Die worden ingezet om familie en vrienden van inzittenden op te vangen . Neergehaald Een adviseur van de Oekraïense regering heeft tegen Interfax gezegd dat alle inzittenden zijn **omgekomen** . Verder zegt hij dat het toestel door pro - Russische **separatisten** met een raket uit de lucht is **gehaald** . President Porosjenko spreekt van een **terroristische daad** . De Europese organisatie voor veiligheid in de Luchtvaart , Eurocontrol , zegt dat het toestel 10 kilometer hoogte **vloog** toen het van de radar **verdween** . Het toestel **vloog** op een hoogte die door de Oekraïense luchtvaart_autoriteiten als veilig was **bestempeld** . Honderden meters lager lag een gebied waar geen burgervliegtuigen mochten komen . Na de **crash** is het lucht_ruim boven Oost - Oekraïne volledig **gesloten** . Buk De **separatisten beschikken** sinds kort over Buk - raketten van Russische makelij , ook bekend als SA11- en SA17-raketten . De **separatisten ontkennen** dat. Zij **zeggen** dat ze niet over de middelen **beschikken** om een toestel dat zo hoog **vliegt** uit de lucht te **schieten** . Getuigen op de grond zeggen dat de wrak_stukken over een groot verspreid liggen . Dat wijst erop dat het toestel in de lucht uit elkaar is **gevallen** of **geschoten** . Zwarte doos **Separatisten** hebben tegen persbureau Interfax **gezegd** dat ze een zwarte doos van het toestel hebben **gevonden** . Leiders van de **separatisten zeiden** eerder al dat die aan Rusland wordt **overgedragen** . De Oekraïense premier zegt dat hij het **neerstorten** van het toestel laat **onderzoeken** en dat zijn troepen niet op luchtdoelen hebben geschoten . De minister van Defensie zegt dat niet vaststaat dat het toestel is **neergehaald** . Onderzoek Een onderzoek naar de oorzaak van een vliegramp wordt in principe gedaan door het land waar de ramp zich voordeed . Zo is dat internationaal afgesproken . Een land dat veel slachtoffers te betreuren heeft , kan de rol van waarnemer krijgen . President Porosjenko zegt dat Nederlandse experts zijn uitgenodigd om aan het **onderzoek** mee te doen . In Nederland is de Onderzoeksraad voor Veiligheid dan de aangewezen instantie . Dat was ook het geval bij het onderzoek naar het neerstorten van een toestel van Afriqiyah Airways bij Tripoli in 2010 . Noodnummers D - reizen zei tegen de NOS dat deze **vlucht** door 25 klanten is **geboekt** . Een deel had Kuala Lumpur als eindbestemming , anderen wilden doorvliegen naar Australië en Nieuw - Zeeland . Drie van hen hebben hun reis op het laatste moment omgeboekt . Voor familieleden stelt D - reizen een noodnummer in : 023 - 5542555 . Het ministerie van Buitenlandse Zaken heeft ook een noodnummer ingesteld : 070 - 3487770 . Familieleden van mensen die **aan boord** waren , kunnen daar terecht met hun vragen terecht . Malaysia Airlines heeft ook een alarmnummer : 0060 - 378841234 . De Franse minister Fabius van Buitenlandse Zaken heeft gezegd dat er ten minste vier Fransen aan boord waren . Mogelijk waren er ook Belgen aan boord . Het Belgische ministerie van Buitenlandse Zaken heeft eveneens een noodnummer ingesteld : 0032 - 25014000 . Boeing Vliegtuigbouwer Boeing heeft een verklaring op zijn website gezet : " Onze gedachten en gebeden zijn bij de mensen **aan boord** van het vliegtuig van Malaysia Airlines dat boven Oekraïne wordt **vermist** , en ook bij hun families en hun geliefden . "

Structured Data

incident type: aircraft shootdown@en (Q6539177)

incident ID: Malaysia Airlines Flight 17@en (Q17374096)

hasPlace: Ukraine

hasActor: victims, Pro-Russian rebels, aircraft, bemanningsleden, 53rd brigade, Oleg Ivannikov, Russia, Vladimir Tsemach, Ukraine, Oleg Poelatov, Sergej Doebinski, Igor Girkin, Leonid Chartsjenko

Selected predicate

Label: Ride_vehicle

Term POS: VERB

Premon: Click here

FrameNet: Click here

Predicate ID: pr23

Frame Relation: type

Frame Element	Role Type	Annotated	Expressed
Theme	Core	true	true
Vehicle	Core	true	true
Source	Core	true	true
Path	Core	true	false
Goal	Core	true	true
Area	Core	true	false

Frame Annotation Tool

http://frame.cltl.nl/annotation

Logged user: **piek**

Back to dashboard

Logout

test_release | aircraft shootdown@ | Malaysia Airlines Flig | nl | 298 doden bij Malays

Select Annotation Type: Frame Annotation

Select Annotation Task: Remove

Select Typicality range:

Save | Clear selection

298 doden bij Malaysia-crash (source)

In Oekraïne is een passagier_s_**vliegtuig** van Malaysia Airlines **neergestort** dat in Amsterdam is **opgestegen** . Malaysia Airlines bevestigt dat **vlucht** MH17 wordt **vermist** . Dat **gebeurde** in de buurt van Donetsk op 50 kilometer van de Russische grens . Malaysia Airlines zegt in een verklaring dat het toestel 283 passagiers en 15 bemanningsleden **aan boord had** . Daar **zitten** veel Nederlanders **bij** , zei minister Opstelten . Hoeveel is nog onduidelijk . Het toestel was om 12.14 uur van Schiphol **opgestegen** . Het zou om 00.10 Nederlandse tijd in Kuala Lumpur **aankomen** . KLM Vlucht MH17 van Malaysia Airlines is een gecombineerde lijndienst van Malaysia Airlines en KLM , die dagelijks op Kuala Lumpur vliegt . Het KLM - vluchtnummer was KL4103 . KLM zegt in een verklaring dat het met leedwezen kennis heeft genomen van " een mogelijk **incident** " met deze **vlucht** . " We staan in contact met Malaysia Airlines om meer informatie te krijgen . " Schiphol heeft een GRIP2-procedure in werking gezet waardoor er hulp_**verleners** naar de luchthaven zijn **gekomen** . Die worden ingezet om familie en vrienden van inzittenden op te vangen . Neergehaald Een adviseur van de Oekraïense regering heeft tegen Interfax gezegd dat alle inzittenden zijn **omgekomen** . Verder zegt hij dat het toestel door pro - Russische **separatisten** met een raket uit de lucht is **gehaald** . President Porosjenko spreekt van een **terroristische daad** . De Europese organisatie voor veiligheid in de Luchtvaart , Eurocontrol , zegt dat het toestel 10 kilometer hoogte **vloog** toen het van de radar **verdween** . Het toestel **vloog** op een hoogte die door de Oekraïense luchtvaart_autoriteiten als veilig was **bestempeld** . Honderden meters lager lag een gebied waar geen burgervliegtuigen mochten komen . Na de **crash** is het lucht_ruim boven Oost - Oekraïne volledig **gesloten** . Buk De **separatisten beschikken** sinds kort over Buk - raketten van Russische makelij , ook bekend als SA11- en SA17-raketten . De **separatisten ontkennen** dat. Zij **zeggen** dat ze niet over de middelen **beschikken** om een toestel dat zo hoog **vliegt** uit de lucht te **schieten** . Getuigen op de grond zeggen dat de wrak_stukken over een groot verspreid liggen . Dat wijst erop dat het toestel in de lucht uit elkaar is **gevallen** of **geschoten** . Zwarte doos **Separatisten** hebben tegen persbureau Interfax **gezegd** dat ze een zwarte doos van het toestel hebben **gevonden** . Leiders van de **separatisten zeiden** eerder al dat die aan Rusland wordt **overgedragen** . De Oekraïense premier zegt dat hij het **neerstorten** van het toestel laat **onderzoeken** en dat zijn troepen niet op luchtdoelen hebben geschoten . De minister van Defensie zegt dat niet vaststaat dat het toestel is **neergehaald** . Onderzoek Een onderzoek naar de oorzaak van een vliegramp wordt in principe gedaan door het land waar de ramp zich voordeed . Zo is dat internationaal afgesproken . Een land dat veel slachtoffers te betreuren heeft , kan de rol van waarnemer krijgen . President Porosjenko zegt dat Nederlandse experts zijn uitgenodigd om aan het **onderzoek** mee te doen . In Nederland is de Onderzoeksraad voor Veiligheid dan de aangewezen instantie . Dat was ook het geval bij het onderzoek naar het neerstorten van een toestel van Afriqiyah Airways bij Tripoli in 2010 . Noodnummers D - reizen zei tegen de NOS dat deze **vlucht** door 25 klanten is **geboekt** . Een deel had Kuala Lumpur als eindbestemming , anderen wilden doorvliegen naar Australië en Nieuw - Zeeland . Drie van hen hebben hun reis op het laatste moment omgeboekt . Voor familieleden stelt D - reizen een noodnummer in : 023 - 5542555 . Het ministerie van Buitenlandse Zaken heeft ook een noodnummer ingesteld : 070 - 3487770 . Familieleden van mensen die **aan boord** waren , kunnen daar terecht met hun vragen terecht . Malaysia Airlines heeft ook een alarmnummer : 0060 - 378841234 . De Franse minister Fabius van Buitenlandse Zaken heeft gezegd dat er ten minste vier Fransen aan boord waren . Mogelijk waren er ook Belgen aan boord . Het Belgische ministerie van Buitenlandse Zaken heeft eveneens een noodnummer ingesteld : 0032 - 25014000 . Boeing Vliegtuigbouwer Boeing heeft een verklaring op zijn website gezet : " Onze gedachten en gebeden zijn bij de mensen **aan boord** van het vliegtuig van Malaysia Airlines dat boven Oekraïne wordt **vermist** , en ook bij hun families en hun geliefden . "

Structured Data

incident type: aircraft shootdown@en (Q6539177)

incident ID: Malaysia Airlines Flight 17@en (Q17374096)

hasPlace: Ukraine

hasActor: victims, Pro-Russian rebels, aircraft, bemanningsleden, 53rd brigade, Oleg Ivannikov, Russia, Vladimir Tsemach, Ukraine, Oleg Poelatov, Sergej Doebinski, Igor Girkin, Leonid Chartsjenko

Selected predicate

Label: Reserving

Term POS: VERB

Premon: Click here

FrameNet: Click here

Predicate ID: pr41

Frame Relation: type

Frame Element	Role Type	Annotated	Expressed
Client	Core	true	true
Services	Core	true	true
Organization	Core	true	true
Scheduled_time	Core	true	false
Booker	Core	true	true

Notes

ID	Sentence	POS	Evokes	Refers to	Relation frame to incident
1	Six Western tourists were **kidnapped** by Al-Faran on 4 July 1995.	verb	**Kidnapping**	the kidnapping as part of Wikidata item Q2026122	the referent is an instance of the frame.
2	In December 1995, the **kidnappers** left a note that they were no longer holding the men hostage.	noun	**Kidnapping**	Hezbollah	the referent is not an instance of the frame
3	Top Hezbollah official Ghaleb Awali was assassinated in a car bomb **attack** in the Dahiya in Beirut in July 2004	noun	**Attack**	the car bombing as part of Wikidata item Q2026122	the referent is an instance of the frame.
4	Israel **can** get to Hezbollah anywhere in Lebanon	verb	**Possibility**	-	there is no referential relation

Table 1: Examples sentences taken from the English Wikipedia page describing the *2006 Hezbollah cross-border raid* (Wikidata identifier Q2026122). The first column indicates the example sentence identifier, the second shows the example sentence, the third provides the part of speech tag of the target word, the fourth which frame the target word evokes, the fifth column provides information about what the target word refers to, and the last column indicates the relationship between the evoked frame and its referent.

ID	Sentence	Evokes	Frame Element
1	*The Tunisian man* who prosecutors say **perpetrated** last month's terrorist attack [...]	**Committing_crime**	PERPETRATOR
2	*he* **ploughed** a truck into a crowded Christmas market	**Impact**	AGENT
3	*Amri* **carried out** the attack	**Intentionally_act**	AGENT
4	*he* **hijacked** a truck	**Piracy**	PERPETRATOR
5	*he* [...] **shot** its Polish driver	**Hit_target**	AGENT
6	*he* [...] **drove** it into the crowded market	**Operate_vehicle**	DRIVER

Table 3: Example sentences taken from reference texts referencing *Anis Amri* (Wikidata identifier Q28052669), the agent of the *2016 Berlin Attack* (Wikidata identifier Q28036573). The first column indicates the identifier, the second column shows the example sentence with the reference to Anis Amri in italics and the frame-evoking predicate in bold, the value of the third column is the evoked frame, and the last column shows which frame element the reference to Anis Amri expresses.

Data release 1.1

- 4 annotators, 4 months, 8 hours a week
- Event types:
  - *mass shooting, aircraft shootdown*
  - *disease outbreak, riot, natural disaster, music festival*
- reference texts: 172 Dutch, 42 English
- 18,960 mentions with 6,066 of 1,973 lexical units, covering 486 frames (avg. 12.5 mentions per frame)
- 393 multi-words, 812 compounds
- 5,068 links to the structured data (instance data)

Annotation results

Inter-annotator-agreement

1. instance-links	
span matches	89.5%
agreement on span matches	89.4%
2. frames	
span matches	73.7%
agreement on span matches	91.9%
similarity in disagreement	0.59
similarity p-value	0.07
3. frames and instance-links	
joint agreement	97.58%
disjoint agreement	89.94%
joint:disjoint ratio	1:2.3
4. frame elements	
agreement incl. unexpressed FE's	69.5%
agreement excl. unexpressed FE's	94.0%

Classical annotation:

- select frame for main predicate: 47% IAA.
- select frame element given a main frame: 79%

Out-of-sentence-relations

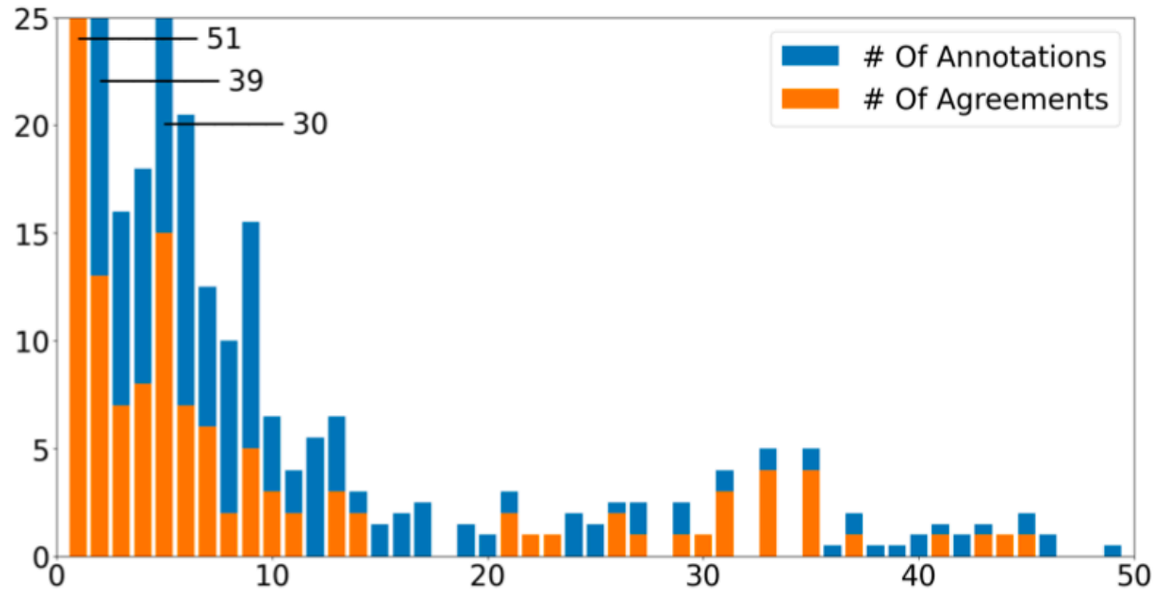


Figure 1: Number of sentence-external frame elements with the distance in sentences to the annotation of the frame. The figure includes the agreement score for each distance to the frame.

26.4% of frame elements outside the frame sentence.

99.8% of frames at least one sentence-external frame element (avg. 1.58 frame element per frame).

Frame Lexicon

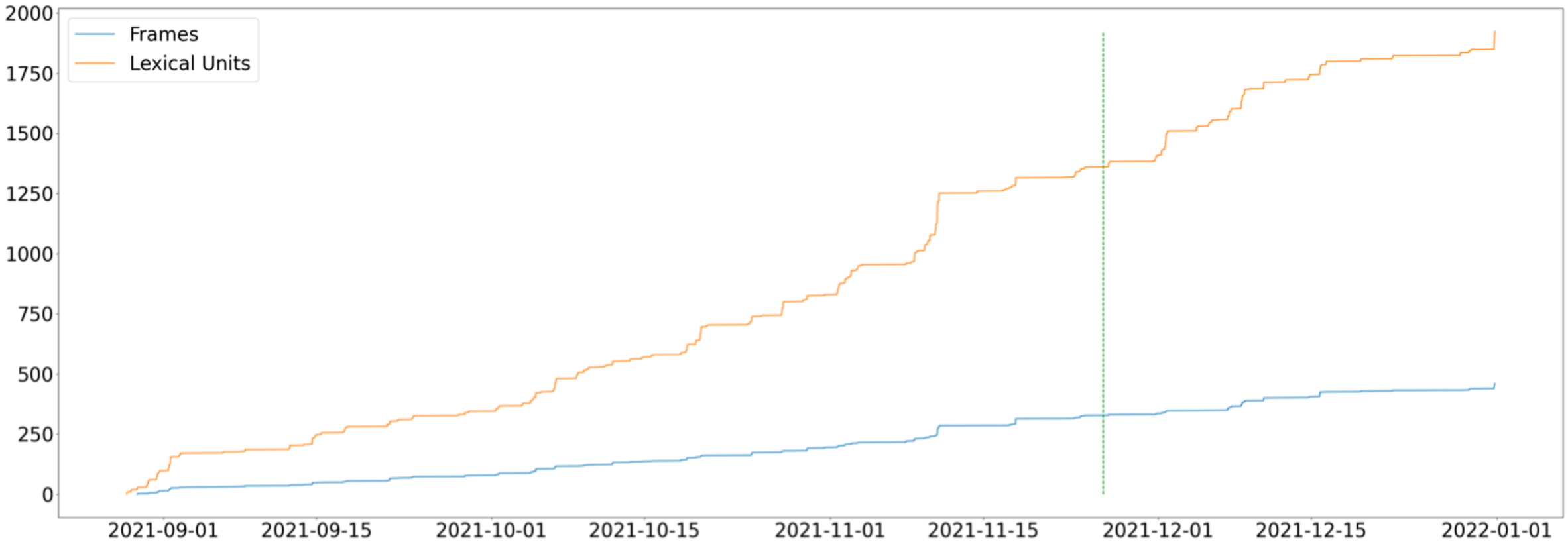


Figure 2: Distribution of new DFN lexicon entries over time, from the beginning to the end of the annotation appointment. The green vertical line indicates the moment that the annotators switched to reference texts of different event types.

Lexical variation for dominant frames

30 fn17-statement ['laten verstaan.V', 'verklaren.V', 'zeggen.V', 'melden.V', 'bekend worden.A', 'spreken.V', 'uitroepen.V', 'bekendmaken.V', 'communiqué.N', 'verklaring.N', 'afkondigen.V', 'beweringen.X', 'te woord staan.X', 'aangeven.V', 'bekend maken.X', 'laten weten.X', 'aankondigen.V', 'twittert.V', 'sprake.N', 'benoemen.V', 'uitspraak.N', 'rapport.N', 'bericht.N', 'suggereren.V', 'uiting.N', 'statement.N', 'making.N', 'verwijzing.N', 'verontschuldiging.N', 'afkondig.V']
23 fn17-killing ['moord.N', 'dodelijk.A', 'bloed.A', 'slachtoffer vallen.N', 'nemen.V', 'murderous.A', 'zelfmoord plegen.V', 'om het leven brengen.V', 'bloedbad.N', 'schietpartij.N', 'zelfmoord.N', 'ombrengen.V', 'wraak.N', 'moorden.V', 'doodmaken.V', 'doden.A', 'uit het leven wegrukken.X', 'liquidatie.N', 'fataal.A', 'een kopje kleiner maken.X', 'suïcidaal.A', 'bloedig.A', 'doodschieten.V']
22 fn17-catastrophe ['ramp.N', 'beset.V', 'X.X', 'kampen.V', 'slachtoffer.N', 'incident.N', 'wrak.N', 'vallen.V', 'tragedie.N', 'drama.N', 'crisis.A', 'victim.N', 'geleden.ADV', 'calamiteiten.A', 'tragisch.A', 'dupe.N', 'noodsituatie.N', 'slachtoffers.N', 'lijden.V', 'overstroming.N', 'wateroverlast.N', 'watersnood.N']
22 fn17-removing ['terugtrekken.V', 'in beslag nemen.X', 'verdrijven.V', 'afvoeren.V', 'intrekken.V', 'afhalen.V', 'halen.V', 'wegtakelen.N', 'verlossen.V', 'verwijderen.V', 'wegrukken.V', 'wegmaken.V', 'ontruiming.N', 'uithalen.V', 'wegdraaien.V', 'wegnemen.V', 'schrappen.V', 'evacueren.V', 'evacuer.N', 'wegspoelen.V', 'afvaren.V', 'evacuatie.N']
19 fn17-causation ['pose.V', 'oorzaak.N', 'push.V', 'brengen.V', 'aanrichten.V', 'toedracht.N', 'aanleiding.N', 'veroorzaken.V', 'aanleiding geven.X', 'maken.V', 'leiden.V', 'ervoor zorgen.X', 'teweegbrengen.V', 'gevolg.X', 'uitslag.N', 'resultaat.N', 'resultaten.N', 'consequentie.N', 'opwekken.V']
19 fn17-emotion_directed ['afschuw.N', 'teleurstelling.N', 'deelneming.N', 'medeleven.N', 'rouw.N', 'hectiek.N', 'consternatie.N', 'meeleven.V', 'verdriet.N', 'belang.N', 'leed.N', 'Condoleance.A', 'condoleance.A', 'jolt.V', 'afgrijzen.V', 'subdue.V', 'blij.A', 'commotie.N', 'ergernis.N']
18 fn17-participation ['meedoen.V', 'deelnemen.V', 'deeldoen.V', 'partij.N', 'betrekken.V', 'betrokkenheid.N', 'betrokkene.N', 'verschijnen.V', 'rol spelen.X', 'spelen.V', 'hand.N', 'de hand hebben.X', 'rol.N', 'te maken hebben.X', 'aandeel.N', 'deelname.N', 'deelnemer.N', 'deelnemend.A']
16 fn17-arriving ['arriveren.V', 'komen.V', 'binnenkomen.V', 'aankomst.N', 'aankomen.V', 'komst.N', 'terugkeer.V', 'benaderen.V', 'bereiken.V', 'terechtkomen.V', 'terugbrengen.N', 'aanrijden.V', 'binnengaan.N', 'terugrijden.N', 'terugkeren.V', 'teruggaan.V']
15 fn17-impact ['geraken.V', 'neerstorten.V', 'storten.V', 'crash.N', 'invloed.N', 'treffen.V', 'inslag.N', 'impact.N', 'crashen.V', 'naar beneden komen.X', 'neerkomen.V', 'schok.N', 'indruk.N', 'raken.V', 'geslagen.V']
15 fn17-judgment ['waardering.N', 'schuld.N', 'aanrekenen.V', 'stellen.V', 'tekortkoming.N', 'wijten.V', 'minachting.N', 'beschuldigen.V', 'op prijs stellen.X', 'accuse.V', 'respecteren.V', 'beoordeling.N', 'eren.V', 'exonerate.V', 'convict.V']
15 fn17-event ['plaats vinden.V', 'situatie.N', 'voorkomen.V', 'plaatsvinden.V', 'houden.V', 'feest.N', 'gebeurtenis.N', 'voordoen.V', 'moment.N', 'plaatshebben.V', 'evenement.N', 'ontwikkeling.N', 'spektakel.N', 'verrijden.V', 'operaties.N']

Lexical variation for dominant frames

```
14 fn17-self_motion ['meelopen.V', 'opmars.N', 'gaan.V', 'dive.V', 'rijden.V', 'stappen.V', 'lopen.V',
             'springen.V', 'rennen.V', 'wandelen.V', 'wuiven.N', 'wegtrekken.V', 'overvlogen.V', 'omslaan.ADV']
14 fn17-performing_arts ['doorspelen.V', 'show.N', 'muzikaal.A', 'artiest.N', 'act.V', 'muziek.N', 'spektakel.N',
             'optreden.N', 'vertolking.N', 'liveshow.N', 'dans.N', 'zanger.N', 'shows.N', 'muzikant.N']
12 fn17-request ['beroep doen op.N', 'oproepen.V', 'eisen.V', 'pleiten.V', 'sommeren.V', 'oproep.N', 'roepen.V',
             'vragen.V', 'verzoeken.V', 'verzoek.N', 'opvragen.ADV', 'uitnodigen.V']
12 fn17-responsibility ['verantwoordelijk.A', 'verantwoordelijkheid.N', 'op hun geweten hebben.X', 'erachter
             zitten.X', 'aansprakelijk.A', 'schuldige.N', 'verantwoordelijkhouden.V', 'verantwoordelijke.N',
             'verantwoording.N', 'achter zitten.X', 'schuldig.A', 'achterzitten.V']
12 fn17-giving ['overdragen.V', 'doorspelen.ADV', 'donatie.N', 'verlener.N', 'geven.V', 'afgeven.V',
             'overhandigen.V', 'bijdragen.V', 'verlening.N', 'weggeven.V', 'gunnen.N', 'inzending.N']
12 fn17-scrutiny ['onderzoeken.V', 'toezicht houden.X', 'opsporing.N', 'toezicht.X', 'doorzoeken.V',
             'zoekingen.N', 'bekijken.V', 'zoeking.N', 'zoektocht.N', 'analyse.N', 'studie.N', 'verdiepen.V']
11 fn17-change_position_on_a_scale ['X.X', 'vallen.V', 'verlagen.V', 'schuiven.V', 'stijgen.V', 'bereiken.V',
             'omhoog.ADV', 'afschalen.V', 'opvaren.V', 'klein maken.X', 'steeg.N']
11 fn17-offenses ['feit.N', 'doodslag.N', 'diefstal.N', 'vernieling.N', 'bespug.N', 'verkrachting.N', 'delict.N',
             'rape.N', 'nalatigheid.N', 'overlast.N', 'overspel.N']
11 fn17-trial ['rechtszaak.N', 'zaak.N', 'rechtszitting.N', 'zitting.N', 'behandeling.N', 'strafzaak.N',
             'terechtstaan.V', 'berechting.N', 'vervolging.N', 'berechten.V', 'go to court.X']
10 fn17-use_firearm ['handwapen.N', 'doorlad.N', 'vuur openen.N', 'gebruiken.V', 'lossen.V', 'het vuur openen.X',
             'raketlancering.N', 'gebruik.N', 'open fire.X', 'vuur.N']
10 fn17-death ['dood.N', 'X.X', 'om het leven komen.V', 'overlijden.V', 'omkomen.V', 'het leven verliezen.X',
             'overleven.V', 'dode.N', 'het leven kosten.X', 'verongelukken.N']
10 fn17-intentionally_act ['actie.N', 'uitvoeren.V', 'vroeg.A', 'maatregel.N', 'handelen.V', 'doen.V',
             'activiteit.N', 'handeling.N', 'maatregel nemen.X', 'voeren.V']
10 fn17-shoot_projectiles ['afvuren.V', 'afschieten.N', 'onder de vuur genomen zijn.X', 'lanceren.V',
             'lancering.N', 'beschieting.N', 'inschieten.V', 'afvoer.V', 'Afvuur.V', 'afschiet.N']
10 fn17-criminal_investigation ['onderzoek.N', 'investigation.A', 'uitzoeken.ADV', 'vervolgen.V', 'Onderzoek.N',
             'onderzoeker.N', 'aanwijzingen.X', 'recherche.A', 'opsporingsonderzoek.N', 'onderzoekers.N']
10 fn17-suspicion ['verdachte.N', 'verdenken.V', 'hoofdverdacht.A', 'verdacht.N', 'verdachten.N',
             'medeverdachte.N', 'hoofdverdachte.N', 'verdachtes.A', 'suspicion.N', 'suspect.V']
```

Most polysemous lexemes

```
5 gaan.V ['fn17-departing', 'fn17-motion', 'fn17-transfer', 'fn17-
        win_prize', 'fn17-self_motion']
3 opnemen.V ['fn17-recording', 'fn17-competition', 'fn17-
        institutionalization']
3 vallen.V ['fn17-change_position_on_a_scale', 'fn17-coincidence',
        'fn17-catastrophe']
3 horen.V ['fn17-hearsay', 'fn17-perception_experience', 'fn17-
        cause_to_perceive']
3 bereiken.V ['fn17-accomplishment', 'fn17-change_position_on_a_scale',
        'fn17-arriving']
3 spelen.V ['fn17-competition', 'fn17-participation', 'fn17-
        performers_and_roles']
```

Framing effects & pragmatic factors

- Granularity of descriptions of situations:
    - Participants described at the level of individuals (people's names), their roles (e.g. suspect or victim), their background (gender, position, race, profession), as members of a group, or simply left out
    - Event as a long-term process, in terms of causes, motivations or intentions and consequences, or physically as sequences of actions.
- Foregrounding and backgrounding and implicit and explicit realisations of frame elements in the text, e.g. focusing on victims and when on suspects
- Historical distance exhibits shifts in perspectives (Cybulska & Vossen, 2010)
- Judgements, hopes and fears, emotions of participants and the expected positive or negative impact of events.

Impact of Historical Distance on framing events

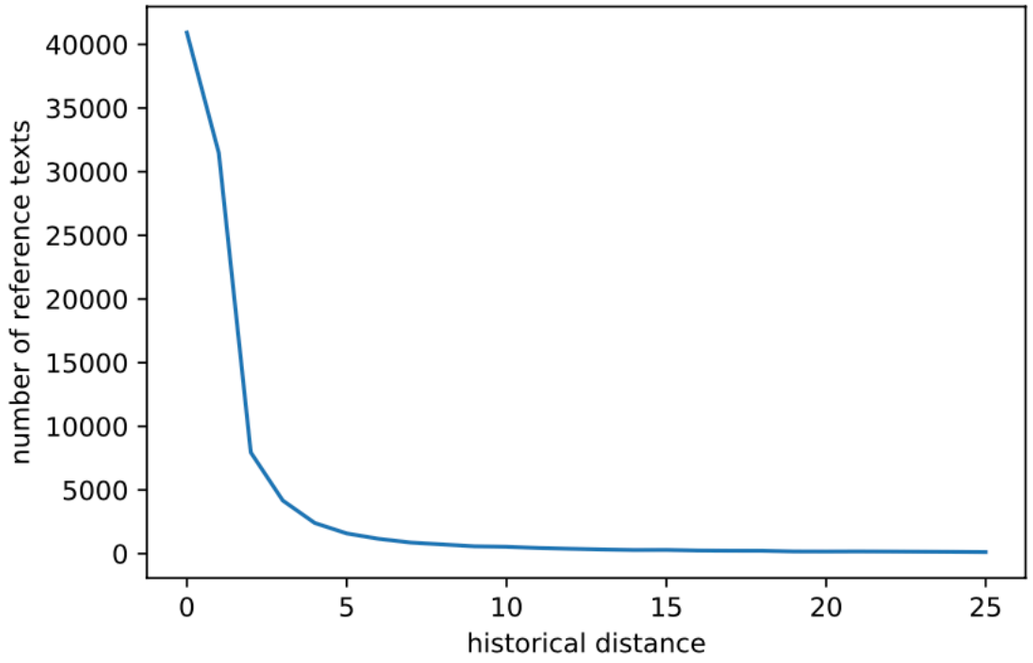


Figure 1: The distribution of the temporal distance is shown for the reference texts that are published within 25 days of the incident, which holds for approximately 90% of the reference texts for the event type gun violence (Q5618454)

Gun violence corpus (James Ko. 2018):
103,090 incidents
123,659 reference texts
Frame annotation using OpenSesame (Yongjie et al 2019)

Gun violence incident

- One man died in a shooting early Thursday morning in southwest Houston. **+1 day**

- One of the four suspects wanted in last week's murder of Keith Thompson was arrested Wednesday morning at a home in Springfield, according to the Jacksonville Sheriff's Office. **+7 days**

Impact of Historical Distance on framing events

rank	day 0	day 8-30
1	STATE_OF_ENTITY (.007566) [D]	JUDICIAL_BODY (.007431) [N]
2	EXPERIENCE_BODILY_HARM (.006752) [Y]	DOCUMENTS (.007431) [N]
3	CAUSE_HARM (.006729) [Y]	JUDGMENT_COMMUNICATION (.006781) [N]
4	EVENT (.006607) [Y]	THEFT (.006538) [D]
5	MEDICAL_CONDITIONS (.006393) [Y]	INTOXICANTS (.006307) [N]
6	TAKING_TIME (.006317) [N]	BAIL_DECISION (.00623) [N]
7	SHOOT_PROJECTILES (.006266) [Y]	ORDINAL_NUMBERS (.006139) [N]
8	DIRECTION (.006037) [D]	CATEGORIZATION (.005915) [N]
9	RESPONSE (.006009) [N]	EVIDENCE (.005842) [N]
10	INFORMATION (.006006) [D]	UNATTRIBUTED_INFORMATION (.005827) [N]
...	...	...
710	KILLING (-.00196) [Y]	KILLING (-.00229) [Y]
711	VEHICLE (-.00299) [D]	VEHICLE (-.00302) [D]
712	LEADERSHIP (-.00422) [D]	CATASTROPHE (-.00421) [Y]
713	ROADWAYS (-.00552) [N]	LEADERSHIP (-.00421) [D]
714	CATASTROPHE (-.005763) [Y]	ROADWAYS (-.00457) [N]
715	AWARENESS (-.00763) [N]	AWARENESS (-.00656) [N]
716	BUILDINGS (-.0093) [Y]	BUILDINGS (-.0072) [Y]
717	LAW_ENFORCEMENT_AGENCY (-.01465) [Y]	LAW_ENFORCEMENT_AGENCY (-.01063) [Y]
718	PEOPLE (-.0379) [Y]	PEOPLE (-.03166)[Y]
719	CALENDRIC_UNIT (-.06463) [Y]	CALENDRIC_UNIT (-.05501) [Y]
720	STATEMENT (-.09892) [N]	STATEMENT (-.08964) [N]

Table 3: The top 10 highest ranked frames (FFICF score)[annotators' score: Y = yes, N = no, D = disagreement] and the 11 bottom ranked frames for the classes "day 0" and "day 8-30" within the event type gun violence. The scores range between -1 and 1.

Conclusions

- Data-to-text PLATFORM to create massive data on situations and their framing across languages

- All code, annotations and lexicons available as open source:

    - http://dutchframenet.nl

- Capture more variation and more representative data

- Enable supervised and unsupervised machine-learning on the data to create semantic role labellers

- Challenges:

    - cover all types of incidents and situations

    - speech-acts, cognitive events

Future plans

- Annotation of more event types
- Fine-tuning models for detection of entities, events, resolve coreference, annotate frames and frame elements, detect subevents
- Pre-annotation of text to be annotated
- Analysis of variation of framing of referentially grounded entities and events across texts in relation to pragmatic factors