# Language as data

#### Piek Vossen Vrije Universiteit Amsterdam

Lorentz workshop: The Future of Academic Lexicography November 4-8, 2019, Leiden

# Language as data

- Empirical approach to language (science, education, technology)
  - Language use defines the problem space rather than theoretical models that should explain it
- How to properly sample language use:
  - data biases & dynamics,
  - capture variation and ambiguity
  - capture the complexity of context
- Annotation is an interpretation of language use —> annotated corpora
- Lexical resources and ontologies are structured abstractions (or generalisations) from these interpretations (annotations)
- Corpora are do not sufficiently capture language as a phenomenon



### Distribution of form and meaning Balance, representativeness and comparability

- DutchSemCor project: <a href="http://wordpress.let.vupr.nl/dutchsemcor/">http://wordpress.let.vupr.nl/dutchsemcor/</a>
  - 25 examples for each meaning of the most polysemous and most frequent words, e.g. "paard" (horse), "band", "spelen" (play), "lopen" (walk)
  - Students had to search for these meanings in the SoNaR corpus of written Dutch (500 million words)
  - 23% of the meanings did not occur in SoNaR
  - but could be found on the Dutch Internet using Webcorp



### Semantic overfitting: form and meaning

MO=mean observed

	Task	Dataset	MOA	MOV	MODA	MODV	EMNLE	ELENM	A=ambiguity
Entity		AIDA test B	1.09	1.35	0.98	0.91	0.05	0.22	V=variation
Linking	EL	WES2015	1.06	1.33	0.97	0.88	0.05	0.21	D=dominance
Entity/		MEANTIME	1.19	4.63	0.98	0.64	0.04	0.55	E-entropy
Classification	EnC	QuizBowl	1.59	1.80	0.92	0.74	0.13	0.46	$p(M_i, L_i)$
<b>_</b> .		ECB	1.61	3.87	0.89	0.61	0.19	0.65	$\frac{p(L_i)}{p(L_i)}$
Event	EvC	ECB+	2.09	3.40	0.85	0.66	0.27	0.57	$p(L; M_i)$
Concretence		TAC KBP '15	4.97	1.22	0.69	0.94	0.47	0.12	$\frac{p(D_i,M_j)}{p(M_j)}$
		SE2 AW	1.20	1.06	0.94	0.98	0.13	0.05	
Word Sense		SE3 task 1	1.21	1.05	0.94	0.98	0.13	0.04	
Disambiguation	WSD	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	
Somantia		SE13 task 12	1.10	1.06	0.97	0.98	0.14	0.05	
Role Labeling	SRL	CoNLL04	1.20	1.00	0.96	1.00	0.09	0.00	

Table 11: Observed ambiguity, variance and dominance.

Ilievski, Postma, and Vossen, Semantic overfitting: what `world' do we consider when evaluating disambiguation of text? COLING 2016.

# Word Sense Disambiguation



- What does machine-learning add to just the MFS statistics?
- How representative are the Senseval/Semeval task? Postma 2019



# **EL Form distribution**

#### llievski 2019

Figure 6: Form frequencies for the instance United\_States.





# EL Meaning distribution

Figure 7: Distributions of the instances denoted by the most ambiguous form ("World Cup").





(b) PageRank distribution.

FIS\_Ski\_Jumping\_World\_Cup 9 1995\_Rugby\_World\_Cup

FIS\_Alpine\_Ski\_World\_Cup

Cricket\_World\_Cup

Rugby\_World\_Cup

FIFA World Cup

1998\_FIFA\_World\_Cup

### What period is covered by our NLP data?



Ilievski, Postma, & Vossen 2016



# 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





# 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



Lorentz workshop: The Future of Academic Lexicography November 4-8, 2019, Leiden

# 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

- Charles Fillmore, The case for case (1968):
  - abstract semantic roles (agent, patient, instrument) in addition to syntactic relations (subject, object)
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. "The berkeley framenet project." 1998.
  - situation dependent roles Perpetrator, Crime, Victim, Offense, Verdict, Sentence
  - words evoke a situational Frame with encyclopaedic knowledge



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



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



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



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



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



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



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



### FrameNet data structure

- Frame: Verdict
- Frame Elements: Case, Charges, Defendant, Finding, Judge
- Lexical Units: verdict.n, ruling.n, pronounce.v, not guilty.a, guilty.a, finding.a, find.v, conviction.n, convict.v, clear.v, acquittal.n, acquit.v
- 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
- **Definition**: A Perpetrator intentionally sets fire to his own or someone else's property, often in order to collect an insurance payment on this property.
- **Example**: Two men were charged with ARSON.



## **Frame relations**





# Frame relations 1









# FrameNet Status

- 1,087 lexicalised frames
- 10,542 frame elements
- 13,640 lexical units in English
- 202,232 sampled(!) annotations in English texts
- FrameNet in other languages: French, Swedish, Japanese, Portuguese, Chinese, German, Spanish, Korean
- Global fragment project: <u>https://www.globalframenet.org</u>



## **FrameNet annotation**



Figure 3.14: Annotation window with target LU  $avenge.\mathbf{v}$ 

- If you have a lexicon, you can lookup the frames and elements
- If NOT, you need to consider all possible relevant frames and elements in each sentence

# Some issues

- Coverage is low: frames, lexical units and annotations are unbalanced or even absent (30% frames, 10% lexical units, 50% of the words lack annotations in test data set, Palmer & Sporleder 2010)
- Frames do not form an ontology and lexical units evoke a frame: arson & arsonist
- Annotation is expensive, slow, incomplete and inconsistent:
  - 20 years to build the English FrameNet AND still not sufficient
  - annotations fit the frames instead of fitting the frames to the texts
- Sentence-based and not Document-based, Syntax-based and not SITUATION-BASED
- There is no FrameNet lexicon for Dutch (!!!)



### **ANNOTATION PROCESS**



- Annotation tool: loads parsed text, provides access to frames through Dutch and English words, provides access to elements after frame is selected
- Validation of disagreements using the visual CAT tool, where we mark all disagreeing annotations and related annotations
- 4 student assistants worked for 6 months (8 hours/week), supervised by a PostDoc



### **ANNOTATION TOOL**

EXPLANATION	
There are several options: (1) Enter the number of the correct frame element. (2) Enter multiple numbers separated by commas if you want to (3) Enter None if none of the roles is the correct one. (4) Enter WrongRelation if there is something wrong with this	o compare some definitions first. 5 particular relation (e.g. this is not an argument of this predicate)
ANNOTATION OF ROLE	
SENTENCE: De vier buitenplaneten stonden toen op een lijn . PREDICATE: stonden	The four outerplanets <b>stood</b> then in one line.
ARGUMENT: De vier buitenplaneten	"The four outer planets were aligned in those days."
YOU HAVE CHOSEN: Being_located	
THE POSSIBLE ROLES FOR THIS FRAME ARE: 0 Theme 1 Place	
2 Dependent_state	
4 Location	
5 Cotheme 6 Depictive	
PLEASE ENTER THE NUMBER(S) OF THE ROLE OF THE ARGUMENT: 0	



### CAT annotation tool

Lenzi, Valentina Bartalesi, Giovanni Moretti, and Rachele Sprugnoli. "CAT: the CELCT Annotation Tool." In LREC, pp. 333-338. 2012.

File Task	Markable Rel	ation Statistics Help	CATool
Corpus 👁		🗸 Confirm Markable 🗙	Delete Selection   📋   Task Selection FN 🗘 / HAS_PARTICIPANT 🗘   💉 Q Q
/Users/antsk	kez/SRLexam kez/SRLexam	S55 Engelsen en Belgen	kennen de bovengisting ( hoge gisting ) , hetgeen resulteert in verschillende biersoorten .
/Users/antsk	kez/SRLexam kez/SRLexam kez/SRLexam	S56 Bovengisting wordt	tegenwoordig ook elders op het vasteland toegepast.
/Users/antsk	kez/SRLexam kez/SRLexam	<b>\$57</b> Veel Belgische bierer	zijn nog steeds bovengistend , en in Nederland bijvoorbeeld de bieren van de Amsterdamse Brouwerij 't IJ , die in Vlaanderen de mosterd haalde.
/Users/antsk /Users/antsk	kez/SRLexam kez/SRLexam	De rond Brussel ges	itueerde lambiekbrouwerijen nemen wat het gistingsproces betreft een aparte plaats in : doordat men de wort een etmaal in open kuipen aan de buitenlucht blootstelt, zorgen in
/Users/antsk	kez/SRLexam kez/SRLexam	de lucht aanwezige g	isten ("wilde gisten") en in eiken tonnen huizende micro-organismen voor een zogenaamde spontane vergisting .
/Users/antsk /Users/antsk	kez/SRLexam kez/SRLexam	S60 Bierwerd waarschij	nlijk al 4500 jaar voor Christus gedronken in Mesopotamië.
FB-FactMark	Relation Attrib	utes	x wen werd bier nog altijd volgens dezelfde oude methode gebrouwen, voornamelijk door vrouwen, en in kloosters en abdijen.
FN-48-Vince	Relation type:	HAS_PARTICIPANT	dijen een zware en een lichte versie van het bier, waarbij de zware versie werd gedronken door paters en gasten, en de lichtere versie door de zusters.
FrameNet2			t brouwen overgenomen door herbergen.
joran-works LREC	frame:	Ingestion	net kruidenmengsels , <mark>ook wel gruit genaamd</mark> .
MASC	sem_role:	ArgM-LOC	Relations
MicroPropos Round1-Uns	frame_element:	Place	S58 petrer < > HAS_PARTICIPANT <>  S58 De rond Brussel gesitueerde lambiekbrouwerijen S58 pemen <>> HAS_PARTICIPANT <>  S58 een anarte plaats
Round1-Uns Round1-Uns Round1-Uns			S58 nemen <-> HAS_PARTICIPANT <-> S58 wat het gistingsproces betreft         S58 blootstelt <-> HAS_PARTICIPANT <-> S58 men
Round1-Uns Round2-Uns			S58 blootstelt <-> HAS_PARTICIPANT <->        S58 een etmaal         S58 blootstelt <-> HAS_PARTICIPANT <->        S58 in open kuipen
Round2-Uns			S58 blootstelt <-> HAS_PARTICIPANT <-> S58 aan de buitenlucht         S58 blootstelt <-> HAS_PARTICIPANT <-> S58 de wort
Markables			S58 zorgen <-> HAS_PARTICIPANT <-> S58 in de lucht aanwezige gisten wilde gisten en in eiken tonnen huizende micro-organismen S58 zorgen <-> HAS_PARTICIPANT <-> S58 voor een zogenaamde spontane vergisting S58 zorgen <-> HAS_PARTICIPANT <-> S58 doordat men de wort een atmaal in open kuinen aan de huitenlucht blootstelt
EVENT_MER *	<sup>c</sup> unsaved values		Soo gedronken <-> HAS_PARTICIPANT <->  Soo uon de la contra la co
		Save	S60 gedronken <-> HAS_PARTICIPANT <-> <-> S60 al 4500 jaar voor Christus S61 gebrouwen <-> HAS_PARTICIPANT <-> <-> S61 volgens dezelfde oude methode
			S61 gebrouwen <-> HAS_PARTICIPANT <-> S61 voornamelijk door vrouwen en in kloosters en abdijen S61 gebrouwen <-> HAS_PARTICIPANT <-> S61 in de vroege Middeleeuwen

### SONAR-KLEIN CORPUS (825K TOKENS)

#### 116 documents with PropBank annotations

theme/genre	nr_of_files	nr_of_annotated_verbs
background-news	3	110
financial	17	1756
medical	1	88
news	5	499
newsletter	3	111
periodicals	37	821
policy	12	352
teletext	3	169
websites	1	49
wiki	34	1295
totals	116	5250



### ANNOTATING DUTCH CORPUS WITH FRAMES AND ELEMENTS

#### 4 student assistants worked for 6 months (8 hours/week), supervised by a PostDoc

Nr. of files	116
Annotated verb tokens	5,250
Annotated verb types	1,335
Unique frames	671
Unique lexical units (lemma-frame combinations)*	4,755
Average nr. lexical units per frame	7

Agreem	Percentage	
Strict frame a	47%	
	hierarchical relations	51%
Lenient frame agreement all relations		54%
Frame element agreem	79%	

\* double annotations, not resolving disagreements



### FROM-DATA-TO-TEXT INSTEAD OF FROM-TEXT-TO-DATA

- Annotating situations in text is expensive, slow and difficult and there is no guarantee that the different text that describe the same situation will get the same annotation
- ► Our solution:
  - Start from a registration of situations
  - ► Find texts that report on the same situation
  - Annotate these texts together given the knowledge of the situation
- Vossen, Piek, Filip Ilievski, Marten Postma, and Roxane Segers. "Don't Annotate, but Validate: a Data-to-Text Method for Capturing Event Data." LREC-2018.



### FROM-TEXT-TO-DATA: HOW TO KNOW THAT TEXTS ARE ABOUT THE SAME EVENT?

### News texts



#### 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,





### FROM-TEXT-TO-DATA: HOW TO KNOW THAT TEXTS ARE ABOUT THE SAME EVENT?

### News texts

**Police Officer Says He Was Fired** 

The Eimes

New Hork (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.

nl.upost .info



Een voormalige Weirton, W.Va., politieagent heeft woensdag een rechtszaak tegen de stad ingediend. Hij beweerde dat hij werd ontslagen omdat hij collega's in gevaar had gebracht nadat hij ervoor gekozen had geen man te schieten en te doden die tijdens een huiselijk geschil een wapen had gezet.

### **Event-Graph**



### FROM-DATA-TO-TEXT FOR A GIVEN TYPE OF EVENT



### FROM-DATA-TO-TEXT

#### ► Advantages:

- Structured information for establishing reference is already given
- Everything that is related to the type of incident needs to be annotated but the set of frames and elements is small and coherent
- Everything that is not related to the type of incident can be ignored (speech acts, cognitive verbs, side-stories)
- ► less conservative annotations, more (true) variation
- ► faster, higher agreement
- better starting point for cross-lingual interoperability



Wikidata Query 🗁 Examples	Help     Image: Tools -     query for crimes	A to English
Image: Output of the second seco	<pre>x     #SELECT (count(distinct ?event) as ?cnt)     SELECT ?event ?eventLabel     WHERE     {         find events             ?event wdt:P31/wdt:P279* wdiQ83267.             # find events             ?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) &amp;&amp; DATATYPE(?date) = xsd:dateTime).             # not in the future, and not more than 31 days ago             BIND(NOW() - ?date AS ?distance).             FILTER(0 &lt;= ?distance &amp;&amp; ?distance &lt; 36500).             # and get a label as well             OPTIONAL {                  ?event rdfs:label ?eventLabel.                  FILTER(LANG(?eventLabel) = "en").</pre>	
		3
Table - 🕜	90 Results in 367 ms	)ownload - & Link -
event	eventLabel	\$
<b>Q</b> wd:Q211087	Pan Am Flight 103	
<b>Q</b> wd:Q217528	Oklahoma City bombing	
<b>Q</b> wd:Q244941	Xi'an Incident	
<b>Q</b> wd:Q429524	2011 Iranian embezzlement scandal	
<b>Q</b> wd:Q460731	Sinking of the Rainbow Warrior	
<b>Q</b> wd:Q536935	1986 West Berlin discotheque bombing	
<b>Q</b> wd:Q569287	Bologna massacre	
Q wd:Q590776	Invergordon Mutiny	





VU



Jones, on Aug. 15, 2002, following years of abuse by the child's mother, Elisha

Jones, the prosecutor said.

VU





#### 7. Reference text original

#### https:// www.volkskrant.nl/ archief/geboren-voorhet-ongeluk~a3307696/

<u>×</u> ~ BANEN TOPICS DIGITALE KRANT SERVICE V **deVolkskrant** Cultuur & Leven **Nieuws** Archief Geboren voor het ongeluk Ton Valkenburg was 'de James Dean van de Bollenstreek'. Bij de grootste luchtvaartramp tot nu toe, 35 jaar geleden op Tenerife, verloor hij vrouw en kinderen. Dorpsgenoot Frénk van der Linden reconstrueert leven en dood van een rijke loser en een charmante ploert met geheimen. 'Wie iets van het menselijk tekort wil begrijpen, moet zich in Ton verdiepen.' FRÈNK VAN DER LINDEN 29 augustus 2012, 00:00  $\square$ 

8. Reference text in the wayback machine

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

INTERNET ARCHIVE https: UayBackMachine 1 capt 12 Oct 2	://www.volkskrant.nl/archief/geboren	n-voor-het-ongeluk~a3307	7696/	Go	SEP <b>4</b> 2016
* *	BANEN	TOPICS	DIGITALE KRANT	SERVICE	WIN
Archiet	vs Cultuur & Leven	deVa	lkski	ant	t
Ge Ton Va	<b>boren voor he</b> t alkenburg was 'de James Dean	t ongeluk van de Bollenstreek'	. Bij de grootste		

luchtvaartramp tot nu toe, 35 jaar geleden op Tenerife, verloor hij vrouw en kinderen. Dorpsgenoot Frénk van der Linden reconstrueert leven en dood van een rijke loser en een charmante ploert met geheimen. 'Wie iets van het menselijk tekort wil begrijpen, moet zich in Ton verdiepen.'

FRÈNK VAN DER LINDEN 29 augustus 2012, 00:00

 $\square$ 

### STRUCTURED EVENT DATA

DBpedia (Knuth et al., 2015; Elbassuoni et al., 2010), Wikidata (Vrandečić and Krötzsch, 2014), and YAGO2 (Hoffart et al., 2013)

event class	wikidata ID	# events
explosion	Q179057	55
crime	Q83267	1.468
natural disaster	Q8065	1.160
accident	Q171558	2.126
sport competition	Q13406554	41.245
election	Q40231	8.178
referendum	Q43109	134
voting	Q189760	17.110
TOTAL		71.476



# Data-to-Text Platform



### Research

- Research questions:
  - Lexical competence and performance of framing situations in Dutch across a variety of genres (inc. historical distance) and for a set of domains:
  - Pragmatic implicatures of using these frames in specific contexts.
- Contrastive corpus analysis on unannotated texts from the different domain texts linked to the types of incidents: e.g. token-type ratios reflecting vocabulary diversity and granularity of locations, time boundaries, and participants of the events.
- Manual (gold-data) and automatic (silver-data) annotations.



# Framing effects

- 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
- Judgements, hopes and fears, emotions of participants and the expected positive or negative impact of events.



### **Extraction Process**





Dublin Core (**dc**): FrameNet (**fn**): Simple Event Model (**sem**): Grounded Annotation FrameWork (**gaf**): Others (**rdfs, rdf, owl time**)

meta data on sources

conceptual situational schema

events, participants, time and location relations anchoring instances to mentions in sources



## Pilot data

- number of incidents: 211
- event types: 175 elections, 36 murders
- total number of text files 633
- Files per language: {'en': 215, 'nl': 212, 'it': 212}
- about <del>12/</del>0.5 hours to generate the data



### Initial statistics on elections

Most common hyperlinks extracted from Wikipedia pages (we only use the first section)

English En	tity	Freq	Dutch Entity	Freq	Italian Entity	Free
Prime Mini	ster	21	parlement	25	Primo ministrare	57
Socialist P	arty	15	vijfde frans republiek	16	Presidente	19
National Assen	nbly	14	chileens	15	Cile	14
Fifth Repu	ıblic	14	zweed	13	Partito Socialdemocratico	14
Presid	dent	14	2006	11	2007	14
Swe	den	13	presidentsverkiezing	10	2011	13
C	hile	12	2011	10	Primo Ministro	13
François Mitterr	and	11	2004	9	Svezia	13
Labour P	arty	10	president	9	2005	12
Riks	dag	10	2005	8	2006	11



# Most dominant frames (using open-sesame)

elect	ion Frame	Freq
Chan	ge_of_leadership	686
	Leadership	443
Interio	or_profile_relation	340
	Calendric_unit	280
	Ordinal_numbers	255
	Time_vector	211
	Aggregate	191
C	ardinal_numbers	183
	Getting	158
	Containing	155



### Validation versus annotation

- Given:
  - the dominant frames and frame elements of an incident
  - explicit references to participants, place and time
- Pre-annotate with frames given the dominant frames
- Validate the pre-annotation
- Add the frame-elements
- Resolve (co-)reference to the same event and entity instances across reference texts in different languages



Frame Annotation	http://fra	aming.fii800.lod.labs.vu.nl	Back to dashboard
Logout			
pilot	murder ᅌ	Q10858948  Coad incident	
	Annotation type:	Choose frame:	Relation type:
	Frame 🗘	Killing	refersTo
			If applicable, pick referents by clicking on the Wikidata links.
	Save annotation	Remove annotation	Clear selection
		incident type: murder	
		incident ID: http://wikidata.org/wiki/Q10858948	
		hasPlace: Russian Empire	
		hasTimeStamp: 1881-03-13T00:002	
		hasActor: Pervomartovtsy, Andrei Zhelyabov, Ignacy Hryni	ewiecki

#### Assassination of Alexander II of Russia . (secondary RT; https://en.wikipedia.org/wiki/Assassination\_of\_Alexander\_II\_of\_Russia)

The assassination of Tsar Alexander II of Russia took place on March 13, 1881 (Old Style : March 1, 1881), in Saint Petersburg, Russia. Alexander was killed while traveling to Mikhailovsky Manège in a closed carriage after one assassing threw a bomb which damaged the carriage, prompting Alexander to dismount, at which point a second assassing threw a bomb that landed at the Tsar 's feet.

Alexander II had previously survived several attempts on his life . The assassination is popularly considered to be one of the most successful actions by the Russian Nihilist movement of the 19th century .

#### The conspirators .

The assassination was planned by the Executive Committee of Narodnaya Volya ( " The People 's Will " ), a terrorist revolutionary organization formed in 1879. Andrei Zhelyabov was the chief organizer. When Zhelyabov was arrested a few days prior to the attack, his wife Sophia Perovskaya took the reins. The assassination.

As he was known to do every Sunday for many years, the emperor went to the Mikhailovsky Manège for the military roll call. He travelled both to and from the " Manège " in a closed carriage accompanied by five Cossacks and Frank (Franciszek) Joseph Jackowski, a Polish noble, with a sixth Cossack sitting on the coachman 's left. The emperor 's carriage was followed by two sleighs carrying, among others, the chief of police and the chief of the emperor 's guards. The route, as always, was via the Catherine Canal and over the Pevchesky Bridge.

The explosion , while killing one of the Cossacks and seriously wounding the driver and people on the sidewalk , had only damaged the bulletproof carriage . The



Loggod upor nick

# Conclusion

- Data-to-text PLATFORM to create massive data on situations and there framing across languages
- All code, annotations and lexicons available as open source:
  - <u>http://dutchframenet.nl</u>
- Capture more variation and more representative data
- 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

