Using Language Learner Data for Metaphor Detection

Egon Stemle (Eurac Research) and
Alexander Onysko (University of Klagenfurt)

Background

- Participation in Metaphor Shared Task Feb/March 2018 (Workshop NAACL 18)
  - Automatic metaphor detection in VU Amsterdam Metaphor Corpus

Outline

- Conceptual metaphor theory [Alex]
- Basis of the shared task: VU Amsterdam Metaphor Corpus [Alex]
- Our approach: hypothesis [Alex / Egon]
- Computational implementation [Egon]
- Results and discussion [Alex / Egon]

Conceptual Metaphor

=> "…metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature."

=> "If we are right in suggesting that our conceptual system is largely metaphorical, then the way we think, what we experience and what we do every day is very much a matter of metaphor."

(Lakoff & Johnson 1980: 3)

Conceptual Metaphor

- Example 1:
  - I don't think this relationship is *going anywhere*.
  - We'll just have to *go our separate ways*.
  - We're at *a crossroads*.
  - This relationship is *a dead-end street*.

=> LOVE (RELATIONSHIP) IS A JOURNEY

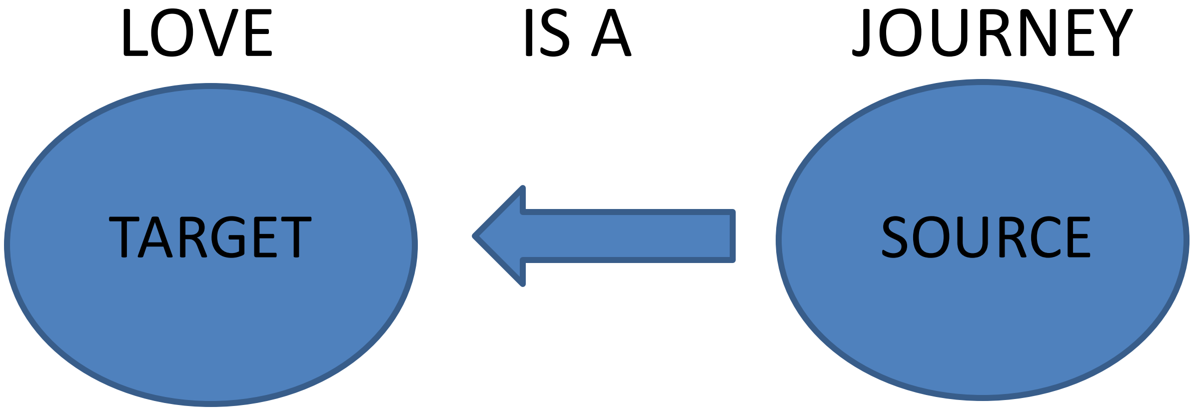
Conceptual Metaphor

- Example 2:
  - Your claims are *indefensible*.
  - His criticisms were *right on target*.
  - She *shot down* all of my arguments.
  - He *destroyed* my opinion.

  => ARGUMENT IS WAR

Conceptual Metaphor

- Structure of conceptual metaphors:
  - Association btw a **target** and a **source**
  - Metaphorical link btw domains consists of correspondences/**mappings**



Conceptual Metaphor

- Cognitive evidence for conceptual metaphor theory:
  - Embodied simulation processes:
    - Experience of physical warmth → more emotional attachment to people (cf. Citron & Goldberg 2014)
    - People moving things up/down while recounting autobiographical memories: up = more positive vs. down = more negative (Casasanto & Dijkstra 2010)

Conceptual Metaphor

- Questions:
  - Level of describing conceptual metaphors (domain labels)?
    - Primary metaphors vs. complex metaphors

      PROGRESS IS MOTION vs. LIFE IS A JOURNEY

  - Universality vs. cultural specificity?

Conceptual Metaphor

- Basic premise
  - Linguistic evidence → conceptual processes (conceptual metaphors)
  - Surface text → metaphorical expressions

Shared Task

=> Goal: to detect, at the word level, all metaphors in a given text.

- Two tracks:
  - All Part-Of-Speech (POS)
    - All content words (nouns, verbs, adverbs, adjectives)
  - Verbs
    - Exclusion of all forms of *be*, *do*, and *have* for both tracks.

Shared Task - dataset

- VU Amsterdam Metaphor Corpus (2010):
  - BNC Baby (text fragments of BNC)
  - 117 text fragments of 4 genres (academic, fiction, news, conversation); ≈ 190,000 lexical units
  - Uneven distribution of metaphors across genres:
    - Academic: 18.5 %
    - News: 16.4 %
    - Fiction: 11.7 %
    - Conversation: 7.7 %

Shared Task - dataset

- VU Amsterdam Metaphor Corpus (2010):

Data	Training			Testing		
	#texts	#tokens	%M	#texts	#tokens	%M
Verbs						
Academic	12	4,903	31%	4	1,259	51%
Conversation	18	4,181	15%	6	2,001	15%
Fiction	11	4,647	25%	3	1,385	20%
News	49	3,509	42%	14	1,228	46%
All POS						
Academic	12	27,669	14%	4	6,076	24%
Conversation	18	11,994	10%	6	5,302	10%
Fiction	11	15,892	16%	3	4,810	14%
News	49	17,056	20%	14	6,008	22%

Verbs and All POS datasets. The table reports the number of text fragments from BNC, number of tokens and percentage of tokens marked as metaphor group by genres.

Source: Leong, Klebanov & Shutova (2018: 58)

VU Amsterdam Metaphor Corpus

- Manual annotation according to MIPVU (cf. Steen et al. 2010):
  - A lexical unit is marked as metaphorical "if the lexical unit has a more basic contemporary meaning in other contexts than the given context" and if "the contextual meaning contrasts with the basic meaning but can be understood in comparison with it."

  Pragglejaz (2007: 3)

VU Amsterdam Metaphor Corpus

- Manual annotation according to MIPVU (cf. Steen et al. 2010):
  - Metaphor related words (mrw):
    - Indirect metaphors
    - Direct metaphors
    - Implicit metaphors (e.g. anaphoric reference - pronouns,...)

VU Amsterdam Metaphor Corpus

- Problematic cases of manual annotation:
  - The 63-year-old **head** of Pembridge Investments, **through** which the bid is being **mounted** says, '**rule [?]** number one **in this [?]** business is: the more luxurious the luncheon rooms at **headquarters**, the more inefficient the business'. [a1efragment01-5]

  - There are other **things** he has, **on** his own **admission [?]**, not fully investigated, like the value of the DRG properties, or which **part** of the DRG business he would **keep after [?]** the *break up **[!]***. [a1efragment01-7]

Our approach: hypothesis

- Using language learner data to train machine learning algorithm (neural networks)
  - Relationship of metaphor use and EFL proficiency (Littlemore et al. 2014; Beigman Klebanov & Flor 2014)

    → more metaphors = higher proficiency

Selection of corpora

- Learner Data:
  - TOEFL 11 (T11) corpus (Blanchard et al. 2013):
    - Test taker essays (proficiency subsets: high, medium, low)
  - VOICE: Corpus of ELF
- Reference corpora:
  - BNC, enTenTen13, ukWaC, ukWaC (T11 size), Wikipedia17

Computational implementation

- Vector Space Model

Words that occur in similar contexts tend to have similar meanings.

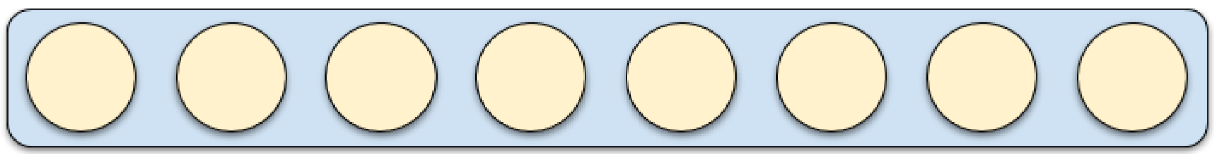
"You shall know a word by the company it keeps"

(Firth, 1957)

(...; Wittgenstein, 1953; Harris, 1954; Weaver, 1955; ...)

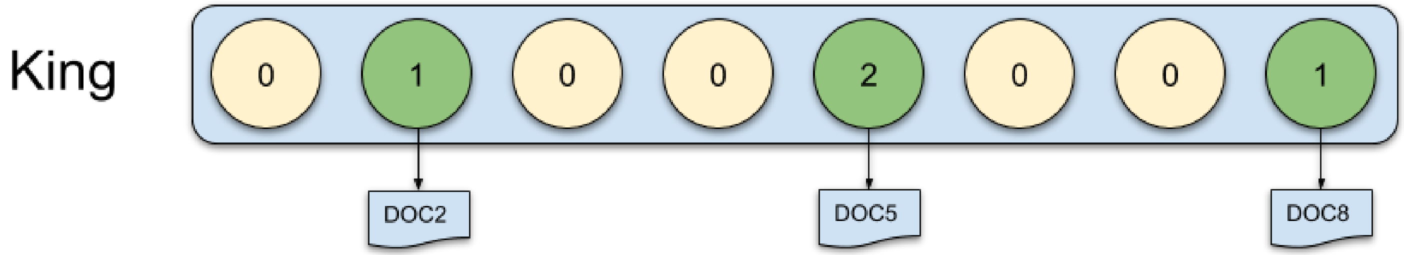
Computational implementation

- Vector Space Model



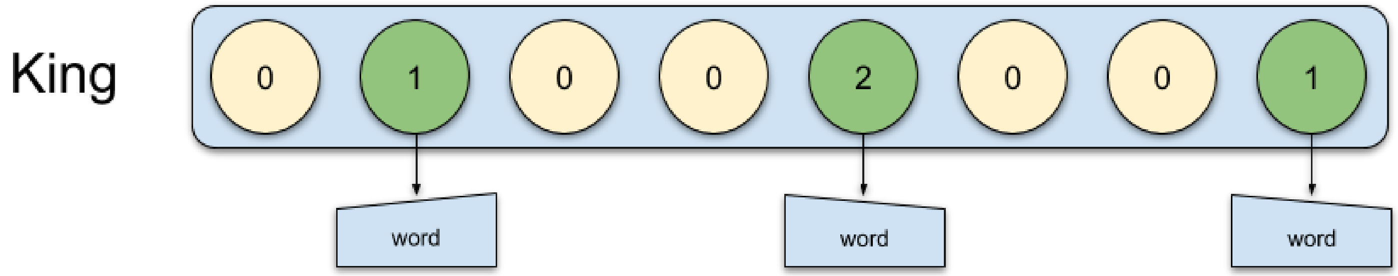
Computational implementation

- Vector Space Model



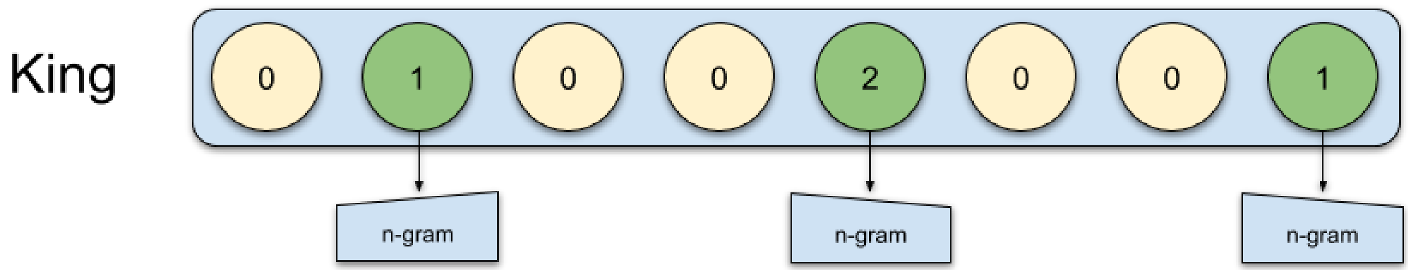
Computational implementation

- Vector Space Model



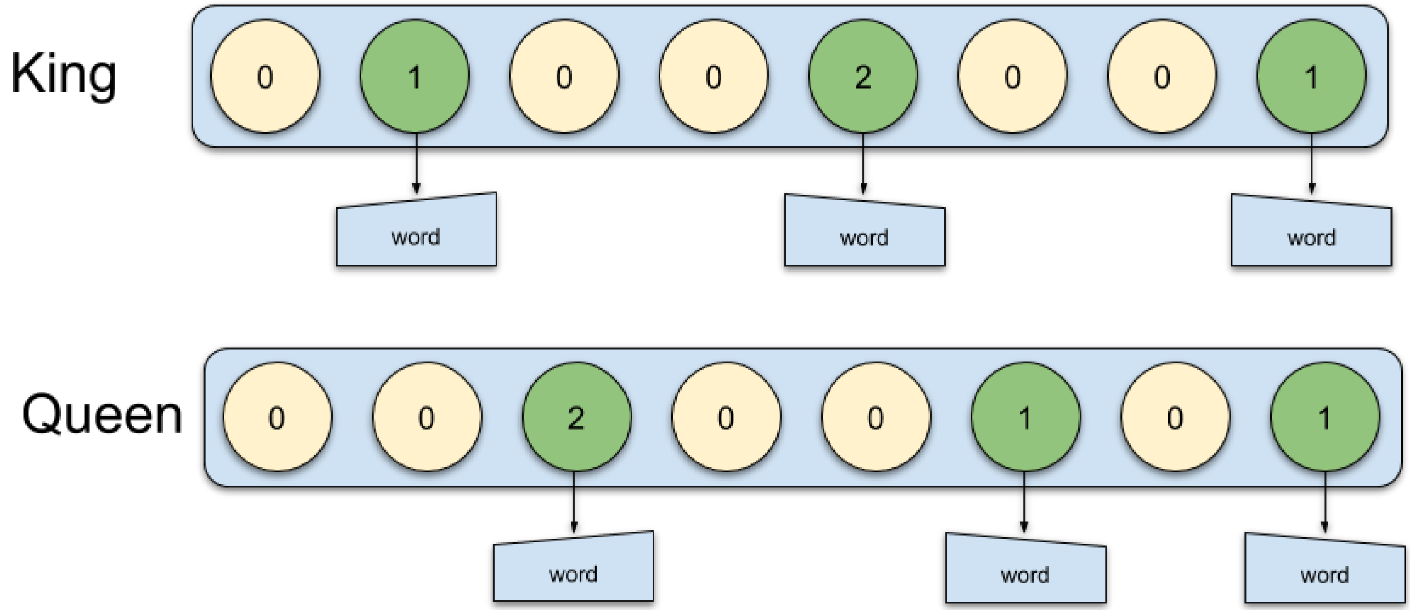
Computational implementation

- Vector Space Model



Computational implementation

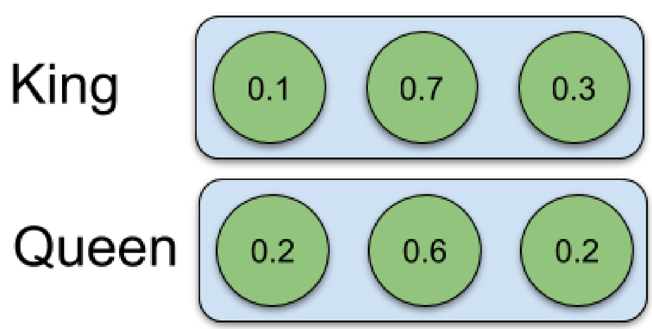
- Vector Space Model



Computational implementation

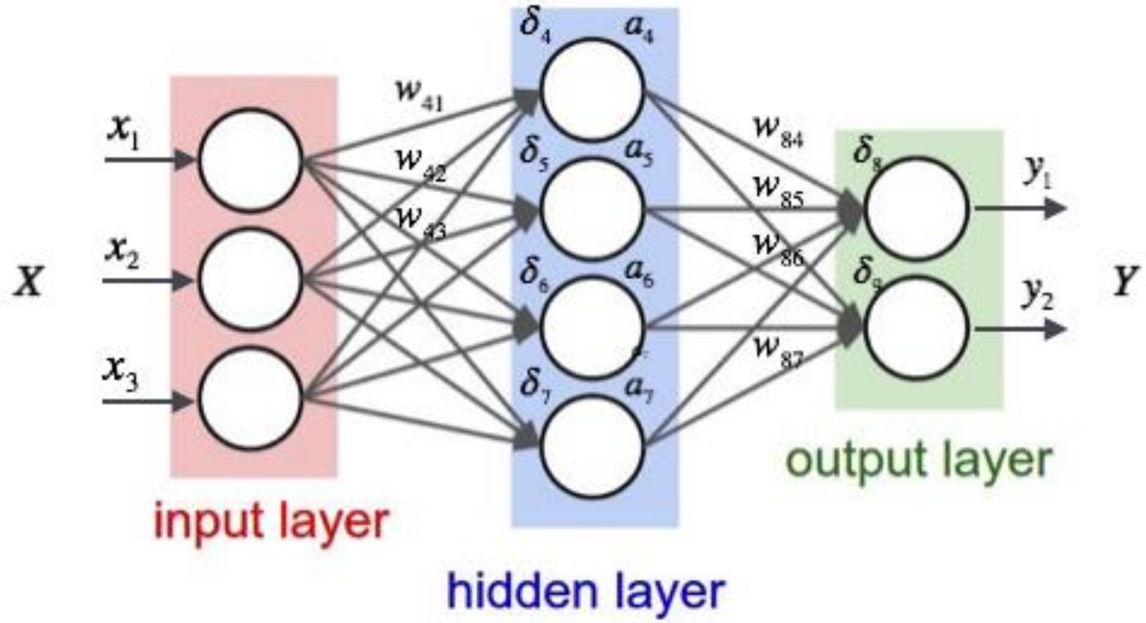
- Vector Space Model, Semantic Vector Space, Distributional Semantic Model, Word Embeddings

PCA/SVD/LSA/LSI/LDA/NN



Computational implementation

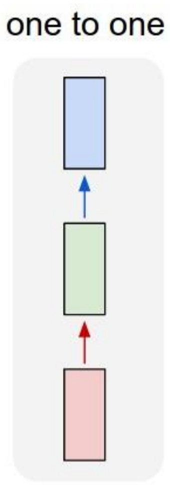
- Neural Networks



Source: medium.com/@curiousily

Computational implementation

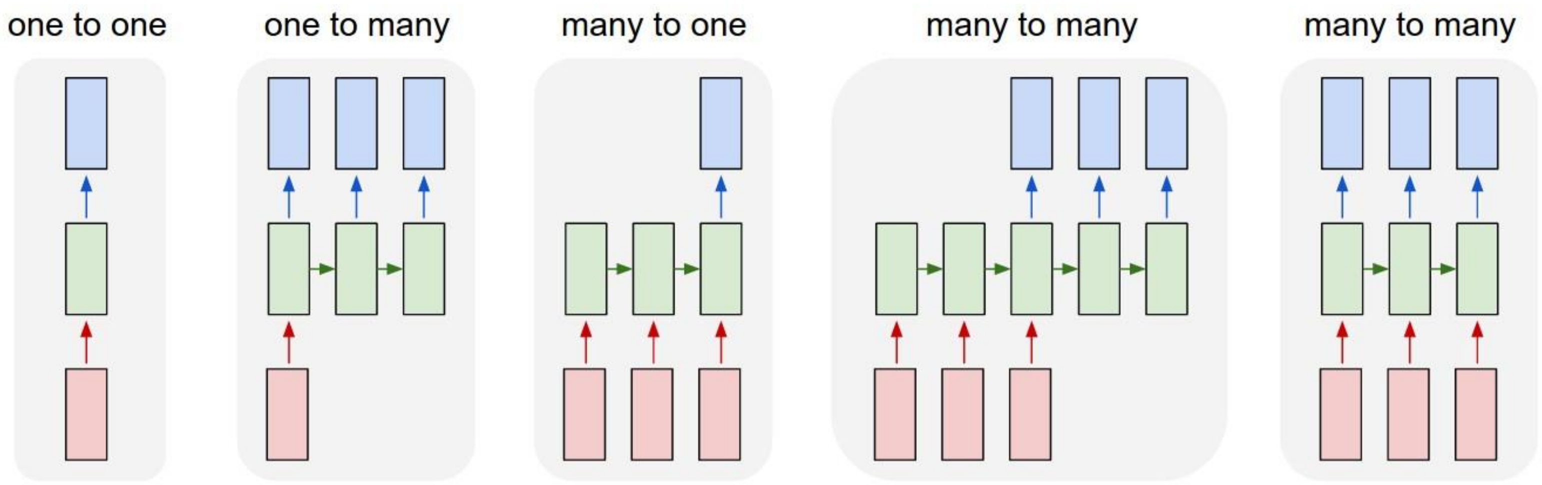
- Recurrent Neural Networks



Source: [Andrej Karpathy blog]

Computational implementation

- Recurrent Neural Networks



Source: [Andrej Karpathy blog]

Computational implementation

Our system combines *trending techniques*, which implement matured methods from NLP and ML.

In particular, word embeddings from standard corpora and from *corpora representing different proficiency levels of language learners* in a LSTM BiRNN architecture.

Shared Task – results (Leong, Klebanov, Shutova 2018: 61-62)

Overall results for all POS

Rank	Team	P	R	F1	Approach
			All POS (Overall)		
1	THU NGN	0.608	0.700	0.651	word embeddings + CNN + Bi-LSTM
2	OCOTA	0.595	0.680	0.635	word embeddings + Bi-LSTM + linguistic
3	bot.zen	0.553	0.698	0.617	word embeddings + LSTM RNN
4	Baseline 2	0.510	0.696	0.589	UL + WordNet + CCDB + Logistic Regression
5	ZIL IPIPAN	0.555	0.615	0.583	dictionary-based vectors + LSTM
6	Baseline 1	0.521	0.657	0.581	UL + Logistic Regression
7	DeepReader	0.511	0.644	0.570	word embeddings + Di-LSTM + linguistic
8	Samsung_RD_PL	0.547	0.575	0.561	word embeddings + CRF + context
9	MAP	0.645	0.459	0.536	word embeddings + Bi-LSTM + CRF
10	nsu_ai	0.183	0.111	0.138	linguistic + CRF

Overall results for verbs

Rank	Team	P	R	F1	Approach
			Verbs (Overall)		
1	THU NGN	0.600	0.763	0.672	word embeddings + CNN + Bi-LSTM
2	bot.zen	0.547	0.779	0.642	word embeddings + LSTM RNN
3	ZIL IPIPAN	0.571	0.676	0.619	dictionary-based vectors + LSTM
4	DeepReader	0.529	0.708	0.605	word embeddings + Di-LSTM + linguistic
5	Baseline 2	0.527	0.698	0.600	UL + WordNet + CCDB + Logistic Regression
6	MAP	0.675	0.517	0.586	word embeddings + Bi-LSTM + CRF
7	Baseline 1	0.510	0.654	0.573	UL + Logistic Regression
8	nsu_ai	0.301	0.207	0.246	linguistic + CRF

Source: Leong, Klebanov, Shutova (2018: 61f.)

Results

	Tokens (Mio)	**min Cnt**	**dim**	T11 (low)	T11 (med)	T11 (high)	T11 (l+m+h)	VOICE	BNC	enTenTen13	ukWaC	ukWaC T11-size	Wikipedia17	F1-score on Test Set	10-fold CV Accuracy on Training Set μ — σ	
T11 (low)	0.3	1	50	X										0.207	0.917	0.016
T11 (med)	1.8	1	50		X									0.526	0.924	0.011
T11 (high)	1.4	1	50			X								0.514	0.930	0.007
T11 (l+m+h)	3.5	1	50				X							0.541	0.928	0.008
VOICE	1	1	50					X						0.495	0.923	0.010
BNC	100	5	100						X					0.597	0.942	0.005
enTenTen13	19,000	5	100							X				0.594	0.947	0.004
ukWaC	2100	5	100								X			**0.598**	0.945	0.004
ukWaC T11-size	3.5	1	50									X		0.564	0.933	0.009
Wikipedia17	ca 2300	5	300										X	0.586	**0.947**	0.003

Results

	Tokens (Mio)	**min Cnt**	**dim**	T11 (low)	T11 (med)	T11 (high)	T11 (l+m+h)	VOICE	BNC	enTenTen13	ukWaC	ukWaC T11-size	Wikipedia17	F1-score on Test Set	10-fold CV Accuracy on Training Set μ — σ	
	7			X	X	X						X		0.576	0.941	0.003
	7						X					X		0.567	0.936	0.008
	103.5			X	X	X			X					0.596	0.944	0.008
	103.5						X		X					**0.613**	0.945	0.005
	103.5								X			X		0.597	0.948	0.003
	104.5			X	X	X		X	X					0.601	0.950	0.004
	107						X		X			X		0.586	0.951	0.002
	108						X	X	X			X		0.550	0.948	0.003
	19,004.5			X	X	X		X		X				0.603	0.947	0.006
	21,400								X	X			X	0.605	0.951	0.003
	21,401							X	X	X			X	0.594	**0.953**	0.003
	21,404.5			X	X	X		X	X	X			X	0.597	0.952	0.003

Results

Genre	F-value	No metaphor (cor / incor)	Metaphor (cor / incor)
Academic	0.696	4150 / 459	1028 / 439
News	0.644	4146 / 541	885 / 436
Fiction	0.505	3574 / 586	418 / 232
Conversation	0.502	4268 / 490	347 / 194

Results

- They held up the bright new diesel buses and, even worse, blocked the progress of private motorists in bulbous Austins and lumpen Humbers in canyon-like city thoroughfares.

- They were inflexible in operation, draughty [!], and mobile [!] reminders of TB epidemics with the enamelled, Do Not Spit, signs…

Conclusion

- Useful results & they tend to support our hypothesis
  - Large unlabeled data set (readily available for many languages)
  - Small-ish manually annotated data set
  - Does not rely on WordNet, VerbNet, concreteness/abstractness information, etc. (used to be the base in previous workshops)
  - Learner Data does carry some viable information for the task

Thank you

Questions / Discussion