Measuring Similarity of Word Meaning in Context with Lexical Substitutes and Translations

Diana McCarthy

Lexical Computing Ltd.

CICLing 2011



- 17 →

Outline

Background

Word Meaning Representation Word Sense Disambiguation Issues

Alternative Word Meaning Annotations

Lexical Substitution (LEXSUB) Cross Lingual Lexical Substitution (CLLS) Usage Similarity (Usim)

Analyses

Conclusions

A ▶

∢ ≣⇒

Word Meaning Representation Word Sense Disambiguation Issues

Manually produced Inventories: e.g. WordNet

match has 9 senses in WordNet including:-

- 1. match, lucifer, friction match (lighter consisting of a thin piece of wood or cardboard tipped with combustible chemical; ignites with friction; "he always carries matches to light his pipe")
- 3. match (a burning piece of wood or cardboard; "if you drop a match in there the whole place will explode")
- 6. catch, match (a person regarded as a good matrimonial prospect)
- 8. couple, mates, match (a pair of people who live together; "a married couple from Chicago")
- 9. match (something that resembles or harmonizes with; "that tie makes a good match with your jacket")

イロン イ部 とくほど くほとう ほ

Word Meaning Representation Word Sense Disambiguation Issues

What is the Right Inventory?

- many believe we need a coarse-grained level for WSD applications [Ide and Wilks, 2006] (though see [Stokoe, 2005])
- but what is the right way to group senses?

WNs#	gloss
1	a young person
2	a human offspring
3	an immature childish person
4	a member of a clan or tribe

Example *child* WordNet

- for MT use parallel corpora if know target languages
- what about summarising, paraphrasing QA, IR, IE?

Word Meaning Representation Word Sense Disambiguation Issues

What is the Right Inventory?

- many believe we need a coarse-grained level for WSD applications [Ide and Wilks, 2006] (though see [Stokoe, 2005])
- but what is the right way to group senses?

xa	imple <i>child</i>	WordNet SENSEVAL-2 gro	up
	WNs#	gloss	
	1	a young person	
	2	a human offspring	
			-

an immature childish person a member of a clan or tribe

Example	child	WordNet	SENSEVAL-2	group	วร
---------	-------	---------	------------	-------	----

- for MT use parallel corpora if know target languages
- what about summarising, paraphrasing QA, IR, IE?

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Distributional Approaches

context	frequency			
	coach	bus	trainer	
take	50	60	10	
teach	30	2	25	
ticket	8	5	0	
match	15	2	6	

・ロン ・四と ・ヨン ・ヨン

æ

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Vector Based Approaches



<ロ> <同> <同> <三>

- ∢ ≣ →

æ

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Vector Based Approaches



・ロト ・回ト ・ヨト

< ∃⇒

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Vector Based Approaches



A B > A B
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

< ∃⇒

æ

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Distributional Approaches: Nearest Neighbours

Output Word: <closest word> <score> <2nd closest > <score>...



_ ∢ ≣ ▶

æ

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Distributional Approaches: Nearest Neighbours

Output Word: <closest word> <score> <2nd closest > <score>... coach: train 0.171 bus 0.166 player 0.149 captain 0.131 car 0.131

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Distributional Approaches: Nearest Neighbours

Output Word: <closest word> <score> <2nd closest > <score>... coach: train 0.171 bus 0.166 player 0.149 captain 0.131 car 0.131 Grouping similar words [Pantel and Lin, 2002]

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense" Residents say militants in a station wagon pulled up , doused the building in gasoline , and struck a match.

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense"

Residents say militants in a station wagon pulled up , doused the building in gasoline , and struck a match.



Image: A math a math

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense"

Residents say militants in a station wagon pulled up , doused the building in gasoline , and struck a match.



match#n#1

Image: A math a math

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense" This is at least 26 weeks by the week in which the approved match with the child is made.

Image: A math a math

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense"

This is at least 26 weeks by the week in which the approved match with the child is made.



< 47 ▶

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense"

This is at least 26 weeks by the week in which the approved match with the child is made.

#9 something that resembles or harmonizes with; "that tie makes a good match with your jacket"



match#n#9

Word Meaning Representation Word Sense Disambiguation Issues

Word Sense Disambiguation (WSD)

Given a word in context, find the best-fitting "sense"

This is at least 26 weeks by the week in which the approved match with the child is made.

#9 something that resembles or harmonizes with; "that tie makes a good match with your jacket" #8 a pair of people who live together; "a married couple from Chicago"



match#n#9 or possibly match#n#8

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Can Current Level of Performance Benefit Applications?

- Enough context: WSD comes out in 'the statistical wash'
- not enough context and can't do anyway
- IR [Stokoe, 2005, Clough and Stevenson, 2004, Schütze and Pederson, 1995] vs [Sanderson, 1994]
- MT [Carpuat and Wu, 2005b, Carpuat and Wu, 2005a] vs [Chan et al., 2007, Carpuat and Wu, 2007]

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Does this Methodology have Cognitive Validity?

- [Kilgarriff, 2006]
 - Word usages often fall between dictionary definitions
 - the distinctions made by lexicographers are not necessarily the ones to make for an application
- [Tuggy, 1993] Word meanings lie on a continuum between ambiguity and vagueness
- [Cruse, 2000] Word meanings don't have discrete boundaries, a more complex *soft* representation is needed

Alternative Word Meaning Annotations Analyses Conclusions Word Meaning Representation Word Sense Disambiguation Issues

Does this Methodology have Cognitive Validity?

- [Hanks, 2000]
 - Computational procedures for distinguishing homographs are desirable and possible, but...
 - they don't get us far enough for text understanding.
 - Checklist theory at best superficial and at worst misleading.
 - Vagueness and redundancy needed for serious natural language processing
- [McCarthy, 2006] Word meanings between others e.g.

・ロト ・回ト ・ヨト ・ヨト

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

_∢≣≯

Outline

Background

Word Meaning Representation Word Sense Disambiguation Issues

Alternative Word Meaning Annotations

Lexical Substitution (LEXSUB) Cross Lingual Lexical Substitution (CLLS) Usage Similarity (Usim)

Analyses

Conclusions

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

Three Datasets

to compare different representations of word meaning in context

- SemEval-2007 Lexical Substitution (LEXSUB) with Roberto Navigli
- SemEval-2010 Cross-Lingual Lexical Substitution (CLLS) with Rada Mihalcea and Ravi Sinha
- Usage Similarity (Usim) (ACL, 2009) and on going ... with Katrin Erk and Nick Gaylord

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

Image: A math a math

Lexical Substitution

Find a replacement word for a target word in context

For example

The ideal preparation would be a light meal about 2-2 1/2 hours pre-match , followed by a warm-up hit and perhaps a top-up with extra fluid before the match.

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

Image: A math a math

Lexical Substitution

Find a replacement word for a target word in context

For example

The ideal preparation would be a light meal about $2-2 \ 1/2$ hours pre-match , followed by a warm-up hit and perhaps a top-up with extra fluid before the game.

LEXSUB Annotation Interface



Please replace the word in bold with a substitute which preserves the meaning of the sentence:

Sentence #671:

The ideal preparation would be a light meal about 2-2 1/2 hours pre-match , followed by a warm-up hit and perhaps top-up with extra fluid before the **match** .

Substituto	dame	OK
Substitute.	guine	OK
	□nil □extra responses □name	🗹 used a dictionar
Target word is part of phrase:		
Comments:		

Reminder: "You are free to consult a dictionary or thesaurus if it helps, but not another person. Please tick the dictionary box if you did consult a dictionary for any of the items for this word"

< previous | next > | summaries | instructions | logout

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへで

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

・ロト ・ 同ト ・ ヨト ・ ヨト

э

Substitutes for *coach* (noun)



Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

・ロト ・同ト ・ヨト ・ヨト

3

Substitutes for *investigator* (noun)



CLLS interface

	_
Start page/ Lexicon Logout	ANNOTATION
LEXICON : SEVERELY.R <u><< Previous</u> Page 3 of 10 (Has been Annontated!) <u>Next >></u> Perhaps the effect of West Nile Virus is sufficient to extinguish endemic birds already severely stressed by habitat losses .	Possible Translations (comma separated) fuertemente, severamente, duramente, exageradamente Nil Name Used a dictionary Target is part of the following phras- multiword / idiom:
	Comment
	Save
	・ ロト ・ 個ト ・ ミト ・ ミト ・ モー ・ つへで

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

・ロト ・回ト ・ヨト

< ∃⇒

2

Some Translations for *severely*



McCarthy CICLING 2011

CLLS example:*stiff*

- 1) Even though it may be able to pump a normal amount of blood out of the ventricles, the <u>stiff</u> heart does not allow as much blood to enter its chambers from the veins.
- 3) One <u>stiff</u> punch would do it.
- 7) In 1968 when originally commissioned to do a cigarstore Indian, he rejected the <u>stiff</u> image of the adorned and phony native and carved "Blue Nose," replica of a Delaware Indian.

S	LEXSUB substitutes	CLLS translations	
1	rigid 4; inelastic 1; firm 1; inflexi-	duro 4; tieso 3; rigido 2; agarro-	
	ble 1	tado 1; entumecido 1	
3	strong 2; firm 2; good 1; solid 1;	duro 4; definitivo 1; severo 1;	
	hard 1	fuerte 1	
7	stern 1; formal 1; firm 1; unrelaxed	duro 2; forzado 2; fijo 1; rigido 1;	
	1; constrained 1; unnatural 1; un-	acartonado 1; insipido 1	
	bending 1		

Rate how similar in meaning the two boldfaced words below are:

This is sentence pair number 9

(1) This more upright position is most easily and affordably achieved through slapping a riser bar on your setup, and only requires you to buy a bar instead of a **bar** and stem.
(2) For twelve hours Livewire will be broadcasting live from the blue **bar** of Union House at UEA in an attempt to raise as much money as possible for a very worthy cause.

(日) (四) (분) (분) (분) 분

- 01: Completely different
- 02: Mostly Different
- 03: Similar
- 04: Very Similar
- o 5: Identical
- Cannot Decide

Click for Full Instructions

-Comment:

Lexical Substitution (lexsub) Cross Lingual Lexical Substitution (clls) Usage Similarity (Usim)

Image: A math a math

Usim example:

1) We study the methods and concepts that each writer uses to defend the cogency of legal, deliberative, or more generally political prudence against explicit or implicit <u>charges</u> that practical thinking is merely a knack or form of cleverness.

2) Eleven CIRA members have been convicted of criminal <u>charges</u> and others are awaiting trial.

Annotator judgments: 2,3,4

Outline

Background

Word Meaning Representation Word Sense Disambiguation Issues

Alternative Word Meaning Annotations

Lexical Substitution (LEXSUB) Cross Lingual Lexical Substitution (CLLS) Usage Similarity (Usim)

Analyses

Conclusions

< A > < 3

-≣->

Questions

- Are paraphrases correlated with translations?
- Are they correlated with judgments of usage similarity?
- Is this always the case for every lemma?
- Does lack of correspondence in these annotations correspond with lower (Usim) IAA for a given lemma?
- Are there other indicators of difficulty in distinguishing sense?

Data for this analysis

- ▶ data common to CLLS, Usim-1 and -2 and LEXSUB
- ▶ 32 lemmas (Usim-1) + 24 lemmas (Usim-2) (4 lemmas in both)
- Usim take the mean judgments (in these slides not transforming Usim judgments to z scores before taking mean. Subtle difference to paper but the conclusions are the same.)
- Overlap in paraphrases and translations calculated:

Correlation between datasets

datasets	ρ	p-value
LEXSUB-CLLS	0.519	< 2.2e-16
$\mathrm{LEXSUB}\text{-}\mathrm{Usim}\text{-}1$	0.576	< 2.2e-16
$\mathrm{LEXSUB}\text{-}\mathrm{Usim}\text{-}\mathrm{2}$	0.724	< 2.2e-16
CLLS -Usim-1	0.531	< 2.2e-16
CLLS-Usim-2	0.624	< 2.2e-16

・ロト ・回ト ・ヨト ・ヨト

æ

Correlation between datasets ... by lemma

	LEXSUB	LEXSUB	CLLS	Usim	Usim
lemma	CLLS	Usim	Usim	MID	IAA
account.n	0.322	0.524	0.488	0.389	0.66
bar.n	0.583	0.624	0.624	0.296	0.35
bright.a	0.402	0.579	0.137	0.553	0.53
call.v	0.708	0.846	0.698	0.178	0.65

・ロン ・四と ・ヨと ・

Correlation between datasets ... by lemma

LEXSUB	LEXSUB	CLLS	Usim	Usim
CLLS	Usim	Usim	rev MID	IAA
throw.v	lead.n	new.a	fresh.a	new.a
neat.a	hard.r	throw.v	raw.a	function.n
work.v	new.a	work.v	strong.a	fresh.a
strong.a	put.v	hard.r	special.a	investigator.n
dismiss.v	fire.v	rude.a	post.n	severely.r
coach.n	rich.a	coach.n	call.v	flat.a
fire.v	execution.n	fire.v	fire.v	fire.v

・ロン ・四と ・ヨと ・

Correlation between datasets ... by lemma

LEXSUB	LEXSUB	CLLS	Usim	Usim
CLLS	Usim	Usim	rev MID	IAA
throw.v	lead.n	new.a	fresh.a	new.a
neat.a	hard.r	throw.v	raw.a	function.n
work.v	new.a	work.v	strong.a	fresh.a
strong.a	put.v	hard.r	special.a	investigator.n
dismiss.v	fire.v	rude.a	post.n	severely.r
coach.n	rich.a	coach.n	call.v	flat.a
fire.v	execution.n	fire.v	fire.v	fire.v
0.424	0.528	0.674	-0.486	

・ロン ・四と ・ヨと ・

Summary

- word meaning annotations using substitutes, translations and similarity judgments
- reflect underlying meanings in context and allow relationships between usages
- annotations of similarity of usage show highly significant correlation to substitutes and translations
- correlation is not evident for all lemmas
- correlation between these annotations by lemma itself correlates with Usim inter-tagger agreement
- proportion of Usim mid scores by lemma is a useful indicator of low inter-tagger agreement and issues with separability of senses

<ロ> (四) (四) (三) (三) (三)

Ongoing and Future Work

- we also have datasets with WordNet annotations: traditional and graded
- datasets available for evaluating different representations of meaning
- ... particularly fully unsupervised
- use one dataset as an approxmation for another (e.g. substitute overlap as indicator of similarity)
- extent that paraphrases and translations can be clustered



Collaboration with Roberto Navigli and Katrin Erk and Nick Gaylord and Rada Mihalcea, Ravi Sinha and Huw McCarthy

- LEXSUB task web site: http://www.informatics.sussex.ac.uk/research/nlp/mccarthy/ task10index.html
- CLLS web site: http://lit.csci.unt.edu/index.php/Semeval_2010
- Usim and graded word sense annotations from websites of Katrin Erk and Diana McCarthy
- data and R code will be available http://www.cicling.org/2011/software/diana

And finally ...



▲□▶ ▲圖▶ ▲理▶ ▲理▶ -

æ



Thank You! Thanks! Cheers! Arigato Gozaimasu! Domo arigato! Dan san! Ohkini!

< ∃⇒

æ

< □ > < □ >

Carpuat, M. and Wu, D. (2005a).

Evaluating the word sense disambiguation performance of statistical machine translation.

In Proceedings of the Second International Joint Conference on Natural Language Processing (IJCNLP), Jeju, Korea. Association for Computational Linguistics.

- Carpuat, M. and Wu, D. (2005b).
 Word sense disambiguation vs. statistical machine translation.
 In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), Ann Arbor, Michigan.
 Association for Computational Linguistics.
- Carpuat, M. and Wu, D. (2007). Improving statistical machine translation using word sense disambiguation.

イロト イポト イヨト イヨト

In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007), pages 61–72, Prague, Czech Republic. Association for Computational Linguistics.

Chan, Y. S., Ng, H. T., and Chiang, D. (2007).
 Word sense disambiguation improves statistical machine translation.

In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, pages 33–40, Prague, Czech Republic. Association for Computational Linguistics.

Clough, P. and Stevenson, M. (2004).
 Evaluating the contribution of eurowordnet and word sense disambiguation to cross-language retrieval.
 In Second International Global WordNet Conference (GWC-2004), pages 97–105.

Cruse, D. A. (2000).

Aspects of the microstructure of word meanings.

In Ravin, Y. and Leacock, C., editors, *Polysemy: Theoretical and Computational Approaches*, pages 30–51. OUP, Oxford, UK.

Hanks, P. (2000).

Do word meanings exist? *Computers and the Humanities. Senseval Special Issue*, 34(1–2):205–215.

Ide, N. and Wilks, Y. (2006).

Making sense about sense.

In Agirre, E. and Edmonds, P., editors, *Word Sense Disambiguation, Algorithms and Applications*, pages 47–73. Springer.

Kilgarriff, A. (2006).

Word senses.

In Agirre, E. and Edmonds, P., editors, *Word Sense Disambiguation, Algorithms and Applications*, pages 29–46. Springer.

McCarthy, D. (2006).

Relating wordnet senses for word sense disambiguation. In Proceedings of the EACL 06 Workshop: Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together, pages 17–24, Trento, Italy.

 Pantel, P. and Lin, D. (2002).
 Discovering word senses from text.
 In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 613–619, Edmonton, Canada.

Sanderson, M. (1994).

A B > A B >

Word sense disambiguation and information retrieval.

In 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 142-151. ACM Press.

Schütze, H. and Pederson, J. O. (1995). Information retrieval based on word senses.

In Proceedings of the Fourth Annual Symposium on Document Analysis and Information Retrieval, pages 161–175, Las Vegas, NV.

- Stokoe, C. (2005).

Differentiating homonymy and polysemy in information retrieval.

In Proceedings of the joint conference on Human Language Technology and Empirical methods in Natural Language Processing, pages 403–410, Vancouver, B.C., Canada.

Tuggy, D. H. (1993).

Ambiguity, polysemy and vagueness.

Cognitive linguistics, 4(2):273–290.



イロト イポト イヨト イヨト

æ