

Addicter: What's Wrong With My Translations?

Dan Zeman, Mark Fishel
Jan Berka, Ondřej Bojar

Charles University in Prague
University of Zurich



Trento, MTM, 6.9.2011

1



Universität
Zürich^{UZH}

Visualizer and Error Labeler

- ADDICTER = **A**utomatic **D**etection and **D**isplay of **C**ommon **T**ranslation **E**Rrors
- Error labeling part (Mark)
- Visualizing part (Dan):
 - View word-aligned corpora
 - Look up corpus examples of a word
 - Look up word occurrences in phrase table
 - Alignment summary of a word
 - Browse test data
 - In addition to the above, also shows auto-detected errors

HTML Visualization

- Cheap interface (from the developers point of view)
- Displayed by your favorite browser
- Words are clickable
 - Links to their own examples
- Alignments shown using tables
 - Simple sentence pairs possibly better using graphics
 - Complex reordering? Graphics not that good.
 - Besides, it would be difficult to show in HTML.

source

deník aktuálně . cz členy nové rady z ČSSD " vyzkoušel " , jak znají notoricky zavedený slan

target

Aktuálně . cz " tested " the Social Democrat members of the new Council in terms of the wel
current coalition partners .

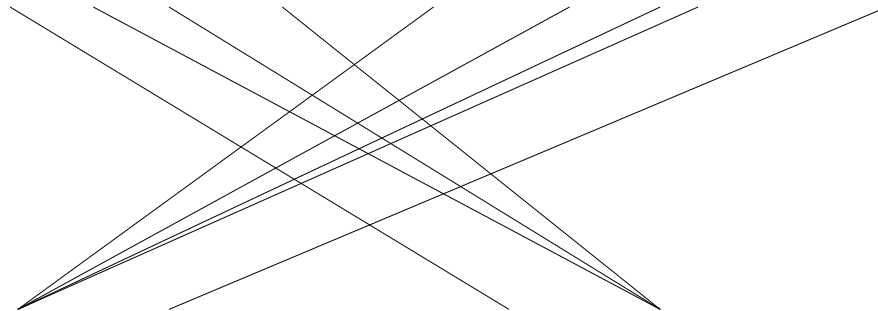
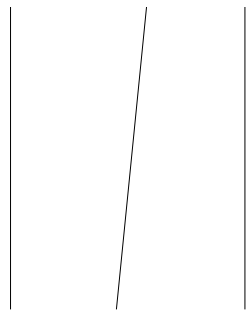
system hypothesis

newspaper currently . cz new Council members from CSSD " " tried , as know notoriously e
current coalition partners .

deník	aktuálně	.	cz	členy	nové	rady	z	ČSSD	"
Aktuálně 0-0 1-0	.	2-1	cz Democrat 3-2 3-8	members of 4-9 4-10	new 5-12	Council 6-13	in terms 7-14 7-15	cz 8-2	" 9-
Aktuálně	.	cz	"	tested	"	the	Social	Democr	
deník aktuálně 0-0 1-0	.	2-1	cz ČSSD vyzkoušel 3-2 8-2 10-2	" 9-3	vyzkoušel 10-4	" 11-5	znají 14-6	cz 3-8	
newspaper	currently	.	cz	new	Council	members	from	CSSD	"
deník 0-0	aktuálně 1-1	.	2-2 3-3	cz nové 5-4	rady 6-5	členy 4-6	z 7-7	ČSSD vyzkoušel 8-8 10-8	" 9

You May Be Used to This...

In the first round, half of the amount is planned to be spent.



V prvním kole bude použita polovina částky.

... or this ...

	V	prvním	kole	bude	použita	polovina	částky.
In							
the							
first							
round,							
half							
of							
the							
amount							
is							
planned							
to							
be							
spent.							

Alignment Summary

The word 'Obama' occurred 1423 times and got aligned to 627 distinct

1. Obama (249)
2. (242)
3. , (56)
4. Obamovi (17)
5. se (16)
6. prezidenta (15)
7. Obamy (14)
8. Obamu (14)
9. prezident (12)
10. že (11)
11. . (11)
12. Obamou (11)
13. a (10)
14. v (9)
15. (7)
16. Barack (6)
17. na (6)
18. k (6)
19. Obamova (6)
20. s (5)

How to Use

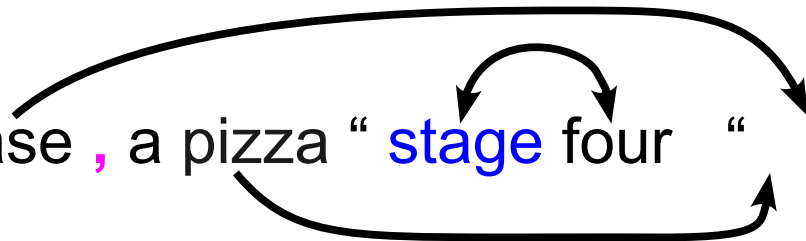
- Word occurrences are first indexed
- Then a Perl script generates the HTML
- Test data browsing: **static HTML**
- Training data / word examples: **dynamic only**
 - Do not pre-generate zillions of pages
 - Drawback: web server + CGI needed

Translation Error Analysis

- Any Single-Number Metric may be good for...
 - comparing two systems **on given dataset**
 - tuning model weights (if easily computable)
- Rarely, if at all...
 - does the absolute value tell anything
- **BUT NEVER...**
 - points directly to the particular **weaknesses** of the system

Error detection and labelling

- src: per favore una pizza “ quattro stagioni “ .
- ref: a “ four seasons “ pizza please .
- hyp-1: one “ four seasons “ pie as a favor .
- hyp-2: please , a pizza “ stage four “ .



Error detection and labelling

- Error taxonomy similar to Vilar et al. (2006)
 - Inflection error / untranslated word
 - Lexical choice error
 - Missing (functional/content word)
 - Superfluous
 - Punctuation
 - Misplaced word (locally/globally)

Error detection and labelling

- Works on word-level
- Requires reference and hypothesis
 - Can benefit from source text, lemmas&PoS-tags
- Uses monolingual alignment
 - Addicter's (...) or any other
 - Requires injective (1-to-1) alignments
 - Can find the “optimal injective subset” for non-injective alignments
- Multiple errors per word allowed

Addicter's alignment

- Lightweight (no learning, no external resources)
- Applied to lemmas (can be done with anything else)
 - Only identical lemmas can be aligned
- HMM-based “disambiguation”
 - $p_{\text{trans}}(a_n | a_{n-1}) \sim \exp(-b * |a_n - a_{n-1} - 1|)$
 - Stimulates to align similarly to previous alignment
 - Exponential time, solved via beam-search

Lexical errors

- Errors are classified, using the alignments:
- Unaligned = missing (in ref) / extra (in hyp)
 - Classified into functional/content via pos-tags
- Aligned: diff. word, same lemma = inflection error
- Aligned: diff. word and lemma = lex. choice error
- Any error on punctuation = punctuation error

Order errors

- To find these, alignment is “unscrambled”
 - Find the minimum number of rearrangements to fix the order
- Transposed adjacent elements = local reordering
- Shifted elements = global reordering

Evaluation

- Data: wmt09 en-cz, 200 sentences * 4 systems
 - Tagged manually with translation errors
- Alignments:
 - Addicter
 - METEOR
 - Bilingual (GIZA++, Berkeley)
 - Via source (CzEng)
- Evaluation: precision/recall of all error tags

Results

Alignment Method	Alignment			Translation Errors		
	Prec.	Rec.	AER	Prec.	Rec.	F-score
addicter&via_source	86.39	85.89	13.86	15.27	54.06	23.82
addicter	98.89	72.18	16.55	10.36	43.76	16.75
addicter&meteor	97.90	71.54	17.33	10.38	43.78	16.78
addicter&giza++intersect	85.99	77.78	18.32	13.47	49.61	21.18
addicter&berkeley&via_source	73.67	83.50	21.72	16.91	54.39	25.80
addicter&berkeley	71.23	78.31	25.40	15.38	52.02	23.74
addicter&giza++grow-diag	65.93	74.58	30.01	14.71	48.56	22.58
via_source	85.00	74.60	20.54	13.80	54.90	22.06
giza++intersect	81.65	64.09	28.19	11.82	48.11	18.97
berkeley*	68.12	74.38	28.89	15.16	51.56	23.43
meteor	90.37	55.04	31.59	6.08	28.68	10.04
giza++grow-diag*	61.54	69.95	34.52	14.50	47.99	22.27

Results

Wrong hypothesis word				Misplaced word			
Flag	Prec.	Rec.	F-score	Flag	Prec.	Rec.	F-score
extra	19.24	64.68	29.65	ows	14.42	48.88	22.27
unk	13.39	12.98	13.18	owl	2.47	47.69	4.70
form	38.16	40.62	39.36	ops	0.00	0.00	0.00
lex/disam	18.48	75.91	29.72	opl	0.00	0.00	0.00
Missing reference word				Punctuation error			
miss_c	2.17	15.28	3.80	punct	29.75	81.65	43.61
miss_a	4.78	27.23	8.14				

Experiment Results

- Underaligned translations => miss/extra overkill
- Dependence on a single reference is bad
- Alignment and error detection quality do not correlate
 - 1-to-1 alignment requirement to blame
 - Have to go to phrase-/syntax-/etc.-based alignments

Future (this week?)

- Lots of improvements possible
- Philipp-style corpus occurrences?, aka collocations
- Index of lemmas
 - Find all occurrences of a word regardless form
- Perl-based web server?
- Further integration between visualization and error analysis
- Further testing of error analysis
- Symbiosis with Hjerson