ToNy: Contextual embeddings for accurate multilingual discourse segmentation of full documents

Philippe Muller, Chloé Braud and Mathieu Morey
philippe.muller@irit.fr, chloe.braud@loria.fr, mathieu@datactivist.coop

Example

[President Bush insists] [it would be a great tool] [for curbing the budget deficit] [and slicing the lard out of government programs.]

President	BeginSeg=Yes
Bush	_
insists	_
it	BeginSeg=Yes
would	_
be	_
a	_
great	_
tool	_
for	BeginSeg=Yes
curbing	_
the	_
budget	_
deficit	_
and	BeginSeg=Yes
slicing	_
the	_
lard	_
out	_
(...)	

References

[Braud et al.(2017a)] Chloé Braud, Ophélie Lacroix, and Anders Søgaard. 2017a. Cross-lingual and cross-domain discourse segmentation of entire documents. In *Proceedings of ACL*.

[Devlin et al.(2018)] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. NAACL 2019 *arXiv:1810.04805*.

[Gardner et al.(2017)] Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke S. Zettlemoyer. 2017. Allennlp: A deep semantic natural language processing platform. arXiv:1803.07640.

Acknowledgements

This work was supported partly by the french PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE.

Problem

- Segmentation is the first step in discourse analysis, crucial but neglected
- Locate arguments or connectives of discourse

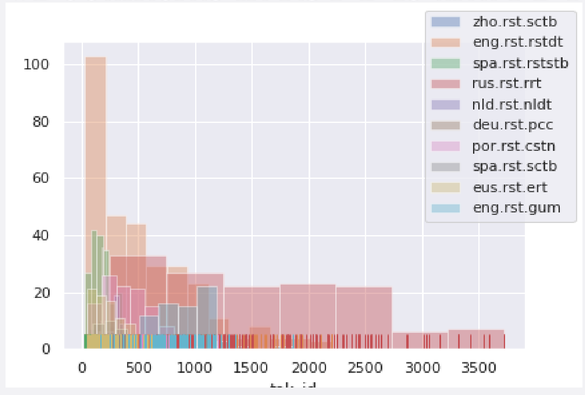
Idea/Hypothesis

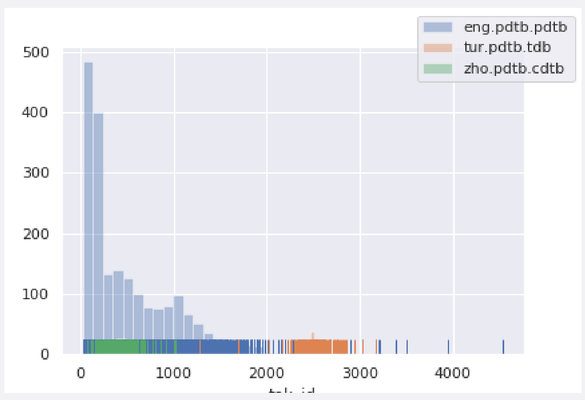
- Segmentation as BIO tagging
- Use only lexical information as given by contextual embedding
- No sentence pre-segmentation necessary ?

DisRPT Shared Task Data

- Multilingual: 9 languages
- Multi-formalism: RST, SDRT, PDTB

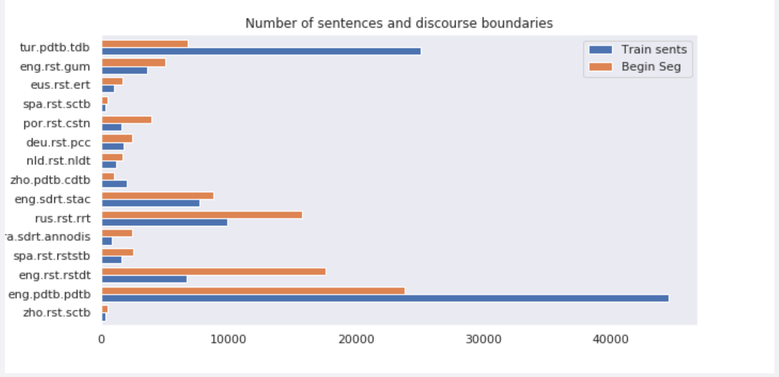
→ #Documents in RST DTs and PDTBs





→

#Discourse Boundaries vs #Sentences



Model/Implementation

- Use AllenNLP library
- Very similar to NER task
- Contextual embedding = BERT, multi-lingual model
- Force segmentation for long text (BERT 512 subword limitation)

Code available at `https://gitlab.inria.fr/andiamo/tony`

Results / PDTB markers

input	corpus	P	R	F1
conll	eng.pdtb.pdtb	89.39	87.84	**88.6**
	tur.pdtb.tdb*	76.89	64	69.85
	zho.pdtb.cdtb	82.67	76.25	**79.32**
	mean	82.98	76.03	79.26
tok	eng.pdtb.pdtb	91.32	87.84	**89.54**
	tur.pdtb.tdb	84.06	86.74	**85.37**
	zho.pdtb.cdtb	81.64	71.07	**75.99**
	mean	85.67	81.88	**83.63**

* with preprocessing error

Results / segmentation

input	corpus	P	R	F1
conll	deu.rst.pcc	95.22	94.76	**94.99**
	eng.rst.gum	95.84	90.74	93.21
	eng.rst.rstdt	95.29	96.81	**96.04**
	eng.sdrt.stac	94.34	96.22	95.27
	eus.rst.ert	89.77	82.87	**86.18**
	fra.sdrt.annodis	94.42	88.12	**91.16**
	nld.rst.nldt	97.9	89.59	93.56
	por.rst.cstn	92.78	93.06	**92.92**
	rus.rst.rrt	86.65	79.49	**82.91**
	spa.rst.rststb	92.03	89.52	**90.74**
	spa.rst.sctb	91.43	76.19	**83.12**
	zho.rst.sctb	87.07	76.19	81.27
	mean	92.73	87.80	**90.11**
tok	deu.rst.pcc	94.88	94.49	**94.68**
	eng.rst.gum	92.28	82.89	**87.33**
	eng.rst.rstdt	93.6	93.27	**93.43**
	eng.sdrt.stac	87.56	80.78	**83.99**
	eus.rst.ert	87.43	80.94	**84.06**
	fra.sdrt.annodis	94.31	89.15	**91.65**
	nld.rst.nldt	94.81	89.97	**92.32**
	por.rst.cstn	93.04	90.72	**91.86**
	rus.rst.rrt	83.37	78.44	**80.83**
	spa.rst.rststb	89.11	90.09	**89.6**
	spa.rst.sctb	87.16	76.79	**81.65**
	zho.rst.sctb	66.26	64.29	65.26
	mean	88.65	84.32	**86.39**

Focus: English discourse annotation

- English has different frameworks: better comparison of models;
- Compare language-specific or multi-lingual approach;
- Compare lexical embeddings: random, contextual or not, BERT/Elmo

	Rand.-50d	GloVe-50d	BERT-E	BERT-M	ELMo
PDTB	77.08	65.17	**90.83**	89.89	88.40
GUM	80.58	78.28	86.29	87.27	**87.65**
RSTDT	78.97	83.21	94.41	93.72	**94.75**
STAC	77.43	71.70	84.65	84.45	**86.06**

RST cross corpus segmentation

Cross corpus comparison with the same framework

- gold sentences;
- dev set;
- single run, different from official eval)

Train/Test	RSTDT	GUM
English RSTDT	93	73
English GUM	66	96