

Is this Sentence Difficult? Do you Agree?

Dominique Brunato⁴, Lorenzo De Mattei⁴ Felice Dell'Orletta⁴, Benedetta Iavarone⁴, Giulia Venturi⁴

 Dipartimento di Informatica, Università di Pisa
 *Scuola Normale Superiore, Pisa
 *Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC–CNR) ItaliaNLP Lab - www.italianlp.it

{dominique.brunato,felice.dellorletta,giulia.venturi}@ilc.cnr.it, lorenzo.demattei@di.unipi.it,beneiavarone@gmail.com



www.italianlp.it

Background and Motivations

Linguistic complexity is a fundamental issue in Linguistics and NLP research. A common distinction is made between an *absolute* notion, which is theory-driven, and a *relative* notion, which is based on the viewpoint of the language user. In the *relative* perspective, linguistic complexity is assessed in terms of (online and offline) processing difficulties resulting from controlled laboratory experiments.

Our Perspective

We approach linguistic complexity in terms of *human perception* as assessed by a judgment of complexity attributed by humans to a given sentence. Unlike traditional studies which typ-

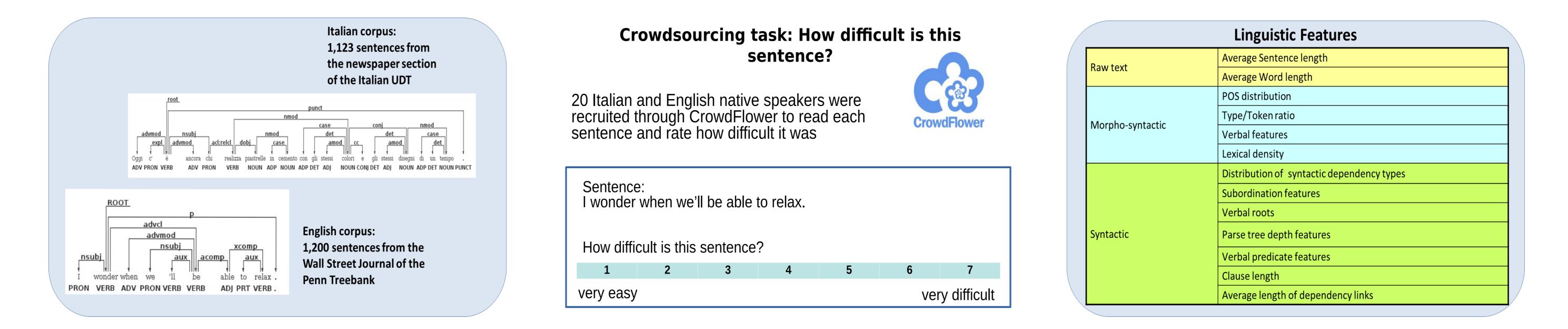
ically assess either lexical or structural complexity phenomena, we focused on the analysis of a wide set of linguistic features to investigate how they all contribute to model how people perceive sentence complexity.

Main Contributions

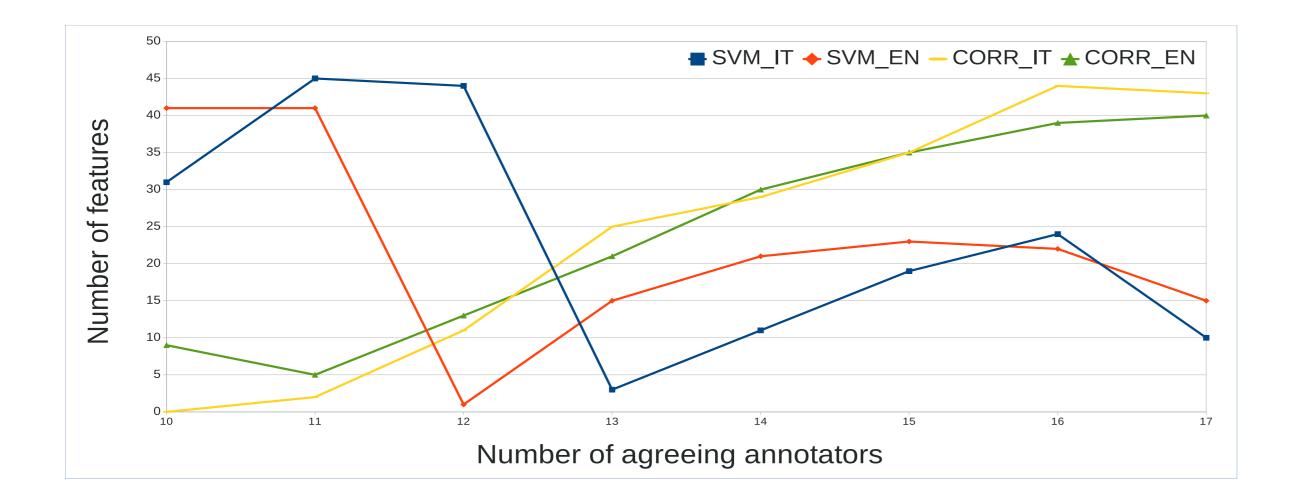
 two research questions aimed to study the role of a set of linguistic phenomena in characterizing a) the agreement among people in assigning the same judgment of complexity to a sentence; b) the human perception of sentence complexity;

a new crowdsourcing-based method to assess how people perceive sentence complexity;
two corpora of sentences annotated by humans with a judgment of complexity.

The Approach



Do You Agree or Not?



- -lexical features, e.g. **type/token ratio**, are significant only for English, while other features related to sentence structure vary only for Italian (i.e. **max depth of the tree**).
- features that are relevant only for the classifier concern the structure and properties of verbal predicate, in terms of morphological features and arity.
- At *higher* degrees of agreement:
- agreed and not agreed sentences in Italian are discriminated by the position of the subject, the position of the object and the presence of subordinate clauses; the presence of an overt subject is significant only for English.

- At *lower* degrees of agreement:
- sentence length and related syntactic features, e.g. the length of the dependency links, vary significantly both for English and Italian.
- features used by the classifier change according to language: e.g. adverbs are used only for Italian, numerals and determiners only for English.
- Baseline and SVM classifier accuracy at different degrees of human agreement:

	Baseline Accuracy (%) – SVM Classifier Accuracy (%)									
	10	11	12	13	14	15	16	17		
Italian	95.4-95.4	91-90.8	80.6-80.5	66.7-66	51.9- 59.1	66.8- 68.8	79- 80.7	87- 87.1		
English	94-94	86.8-86.8	83.6-77.4	66.3-66.1	53.9- 60	60.7- 71.8	70.9- 79.3	80.4- 84.6		

Human Perception of Sentence Complexity

Italian

	L10	L15	L20	L25	L30	L35	TOT10	TOT14
char_tok -	1	1	1	1	1	5	30	31
cpos_ADJ -		7	4	5	7	17	28	23
cpos_ADP -							48	47
cpos_ADV -						22	47	51
cpos_AUX -	-5		-6	-4	-5			
cpos_CONJ -							33	34
cpos_DET -							-6	-8
cpos_NOUN -							-9	
cpos_NUM -	-2		-7	-1	-2	-4	–10	
cpos_PRON -					11		35	41
cpos_PROPN -						-9	_	
cpos_PUNCT -		0		0		-8	-7	-6
cpos_SCONJ -		6		6		6	29	30
cpos_VERB -						9	51 31	37
- dep_acl - dep_acl:relcl							21	20
dep_advcl -						11	18	19
dep_advmod -						20	26	28
dep_adviniod -		2	2	4	9	13	20	18
dep_amou dep_appos -		-			3	10	41	40
dep_appos dep_aux -	-3	-1	_4	-3			45	48
dep_auxpass -				U			49	42
dep_case -			-5			-6	5	2
dep_cc -							17	21
dep_ccomp -						8	38	35
dep_conj -							22	26
dep_det -							9	6
dep_det:poss -					5		40	38
dep_dobj -			9				19	17
dep_expl -							39	43
dep_foreign -					4		50	
dep_mark -						1	16	16
dep_name -							43	49
dep_neg -							46	50
dep_nmod -						-3	4	4
dep_nsubj -							23	29
dep_nsubjpass -				-7	_		52	44
dep_nummod -	_1			-2	-3	-1		
dep_punct -		-		_		-7	13	13
dep_root -		-2		-5		-2	-1	-1
dep_vocative -							-8	-7
dep_xcomp -						-	34	32
links_len -		5				-5	11 1	12
max_depth -		Э				-10		1
- max_links_l - n_prep_chains	3		-3			-10	-5 3	_4 3
n_principal_clauses -	3	<u> </u>	-3		_4		4	<u> </u>
n_subord_chain -					-4	14	-4	9
n_tokens -		4		2		14	2	5
order_obj -			7	_		13	27	22
order_subj -							36	36
order_subord -						15	12	11
prep_chain_l -							32	27
prep_depth -	2		-1				14	14
subord_chain_l -					3	10	10	10
subord_clauses -					2	3	6	7
subord_depth -		3	3		6	23	15	15
token_clause -							37	33
ttr_form -						7	-2	-2
ttr_lemma -						18	-3	-3
verbal_head_arity -	-4		8	-6	8	2	20	25
verb_arity -			-2		-1		53	46
verb_head -						4	7	8
verbs_mood -				3		21	42	45
verbs_num_pers -			6		10	12	25	24
verbs_tense -			5			16	44	39

English

	L10	L15	L20	L25	L30	L35	TOT10	TOT14
char_tok -			3	2			39	42
cpos		6	6				-10	-12
cpos_ADJ -							34	36
cpos_ADP -							40	40
cpos_CONJ -			-24				32	30
cpos_DET -		-3	-5			-4	-7	-7
cpos_NOUN -		0	1	0	4	0	47	43
cpos_NUM - cpos_PRON -	-1	2 _4	-10	3	1 _7	2 _9	17	16
cpos_PRT -	-1	4 5	-10	-2	-2	-9	49	
cpos_VERB -		_5	-1	-7	-3	-2	-6	-5
dep_acomp -			_14	-	J	_	Ū	Ū
dep_adp -							38	38
dep_adpcomp -							41	39
dep_adpmod -							2	2
dep_adpobj -			5				45	44
dep_advcl -							35	41
dep_advmod -			10				30	33
dep_amod -					4		6	8
dep_appos -		7					33	32
dep_attr -		-7 -2	_4	-1	-1	_1	-11 -12	_9 _8
dep_aux - dep_auxpass -		-2	-4 -25	-1 -9	-1	-1	-12	-0
dep_adxpass - dep_cc -			-23				14	15
dep_ccomp -			-21				24	28
dep_compmod -		5	7		6		8	9
dep_conj -			-20				13	13
dep_dep -			8				28	29
dep_det -		1	-6			-6	20	14
dep_dobj -							18	22
dep_expl -			-17					
dep_infmod -				-4			48	47
dep_iobj -								48
dep_mark -			-22				29	31
dep_mwe -		4.4	10			7	46	45
dep_neg -		-11	-18		5	-7	22	21
dep_nmod - dep_nsubj -			-9		5	-10	22	23
dep_nsubjpass -			- <u>-</u> 26			-10	21	20
dep_num -		3	4		3	1	10	7
dep_p -			12			_	5	4
dep_parataxis -					7			
dep_partmod -				1			26	27
dep_poss -		-10					27	25
dep_prt -					-5			
dep_rcmod -					-8		31	26
dep_rel -							42	
dep_ROOT -							-2	-2
dep_xcomp -		-8	-12	-3	<u> </u>		43	46
lex_density -		1	2 11		2	3	36	37
links_len - max_depth -			_19			-12	4 7	5 6
max_deptn - max_links_l -			-19			-12	-5	6
n_prep_chains -							3	3
_principal_clauses -							-4	_4
n_subord_chain -			-11			-8	12	11
n_tokens -							1	1
order_obj -			9				37	34
order_subj -			-3	-5	-10		-9	-10
order_subord -			-16				19	19
prep_depth -		4					9	10
prep_len -			•		^	_	23	20
subord_clauses -		-12	-8		-9	-5	11	12
subord_depth -			-13	<u> </u>	8		16	17
token_clause -			-7	-6	-6		44	35
- ttr_form - ttr_lemma							3 1	_3 _1
verb_arity -		-9				4	-1 -8	-11
verb_head -		U U	-2	-8	-4	-3	15	18
verb_head_arity -		-6	_15	4		-11	25	24
/ _								

- the correlation between the top 20 ranked features and the complexity judgment is extremely high (from 0.30 to 0.85) for sentences at agreement 14 in both languages;
- long sentences were judged as more complex for both languages;
 - at all lengths, sentences were always rated as more complex for Italian;

• at all lengths:

- highly correlated features concern also deep syntactic features, e.g. depth of the whole parse tree, length of dependency links, features related to subordination and nominal modification;
- the two languages differ in terms of:
- language-specific features correlating with complexity, e.g. for English, the distribution of numbers and, for Italian, verbal morphology.
- position in the ranking of features across bins of same-length sentences, i.e. for English the majority of features are similarly ranked in all bins while for Italian rankings differ for sentences < and >20 token long.

Ranking of features correlated with complexity judgments

Predicting Human Complexity Judgments

Performance of a linear SVM regression model in predicting human complexity judgments in terms of i) *mean absolute error* to predict the same complexity judgment assigned by humans and ii) *Spearman correlation* between the ranking of features produced by the model and that by humans.

	IT-10	IT-14	EN-10	EN-14
mean abs err 1	0.77	0.78	0.71	0.68
Spearman 1	0.57	0.64	0.68	0.64
mean abs err 2	0.79	0.80	0.70	0.70
Spearman 2	0.55	0.63	0.67	0.73
mean abs err 3	0.85	0.75	0.77	0.60
Spearman 3	0.55	0.64	0.61	0.71
avg mean abs err	0.80	0.78	0.72	0.66
avg Spearman	0.56	0.63	0.65	0.69

2018 Empirical Methods in Natural Language Processing (EMNLP), October 31–November 4, Brussels.