Is this Sentence Difficult? Do you Agree?

Dominique Brunato[◇], Lorenzo De Mattei[◇•]
Felice Dell'Orletta[◇], Benedetta Iavarone[◇⋆], Giulia Venturi[◇]

[•]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

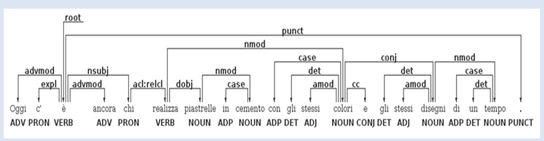
We approach linguistic complexity in terms of *human perception* as assessed by a judgment of complexity attributed by humans to a given sentence. Unlike traditional studies which typically assess either lexical or structural complexity phenomena, we focused on the analysis of a wide set of linguistic features to investigate how they all contribute to model how people perceive sentence complexity.

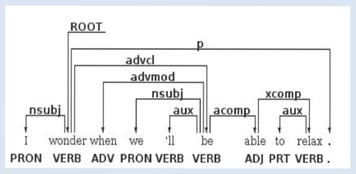
Main Contributions

- 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

Italian corpus:
1,123 sentences from the newspaper section of the Italian UDT





English corpus:
1,200 sentences from the Wall Street Journal of the Penn Treebank

Crowdsourcing task: How difficult is this sentence?

20 Italian and English native speakers were recruited through CrowdFlower to read each sentence and rate how difficult it was

CrowdFlower

Sentence:
I wonder when we'll be able to relax.

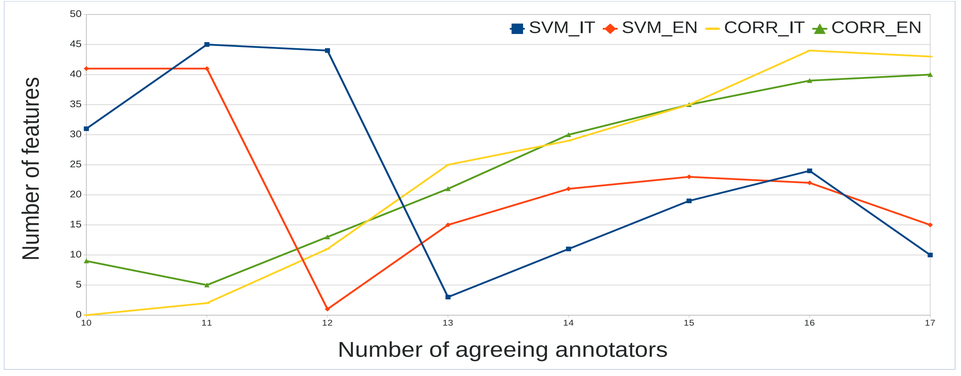
How difficult is this sentence?

1	2	3	4	5	6	7

very easy — very difficult

Linguistic Features

Raw text	Average Sentence length
	Average Word length
Morpho-syntactic	POS distribution
	Type/Token ratio
	Verbal features
	Lexical density
Syntactic	Distribution of syntactic dependency types
	Subordination features
	Verbal roots
	Parse tree depth features
	Verbal predicate features
	Clause length
	Average length of dependency links

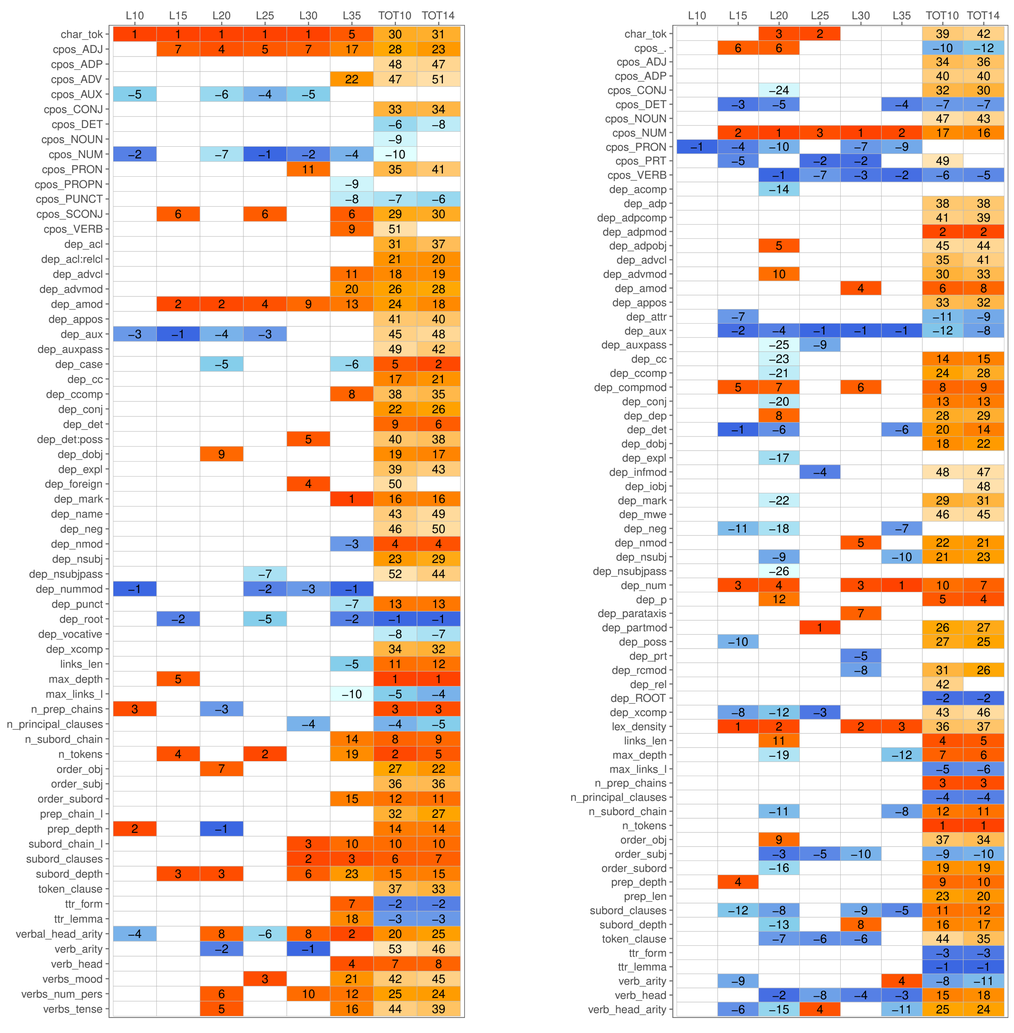


- At *lower* degrees of agreement:
  - **sentence length** and related syntactic features, e.g. the **length of the dependency links**, vary significantly both for English and Italian.
  - lexical features, e.g. **type/token ratio**, are significant only for English, while other features related to sentence 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.
  - 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 ***English***



Ranking of features correlated with complexity judgments

- 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 $\leq$ and $\geq$20 token long.

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**

Do You Agree or Not?

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