Towards FraCaS-BR

Valeria de Paiva[1][0000−0002−1078−6970] and Livy Real[2,3]

[1] Topos Institute, Berkeley, USA
valeria@topos.institute
[2] americanas s.a. d.lab – São Paulo, Brazil
livyreal@gmail.com

The European Commission project "A Framework for Computational Semantics", abbreviated as FraCaS[3], ran from 1993–1996. Its goal of presenting "an informal framework which allows comparison of current semantic approaches both with respect to their claims and their usefulness for implementation" has been very successful. The collection of examples devised to compare the different formalisms is still used, almost thirty years later, as a benchmark for the kinds of reasoning with Natural Language that automated systems should be able to emulate.

The automated systems doing natural language semantics have changed considerably along these decades. The datasets that are used to compare systems have changed considerably too: today's huge datasets, such as SNLI [2] or MultiSNLI [9] have been created to train and evaluate machine learning systems. The machine learning systems created to solve the task of 'Recognizing Textual Entailment' (RTE) –also called NLI (Natural Language Inference)– instead of comparing symbolic (logical) representations of the sentences compare (what we hope are) results of inferences from the sentences.

There are plenty of resources, corpora, scripts, competitions to evaluate NLI in English. There are fewer resources for NLI in Portuguese. The corpus ASSIN [4] (for *Avaliação de Similaridade Semântica e Inferência Textual*) is described in the site HuggingFace[3] as "a corpus annotated with pairs of sentences written in Portuguese that is suitable for the exploration of textual entailment and paraphrasing classifiers." The corpus was used for a shared task as part of the conference PROPOR2016, but unlike most of the big NLI corpora, it distinguishes entailment, paraphrase and neutrality. So it does not consider contradictions, a serious issue.

A second corpus for NLI in Portuguese is SICK-BR [4], a careful translation of the SICK corpus of Marelli et al[5]. Unlike ASSIN the corpus SICK was constructed to be a simple corpus, as far as linguistic phenomena are concerned. Thus named entities, complicated time expressions and world knowledge are kept at a minimum, as is vocabulary. The construction of the corpus and the decisions taken during that process are discussed in [8]. Some of the work that went into the creation of the corpus SICK-BR was used in the second edition of the ASSIN shared task (ASSIN 2), which was collocated with STIL 2019. The task, the corpus used for the task and the evaluation results are described in [7].

---

[3] https://huggingface.co/datasets/assin
[4] https://github.com/livyreal/SICK-BR

The corpus we describe here, FRACAS-BR, goes back to the original goal of benchmarking a large selection of semantic phenomena. This time we use the English FraCaS corpus as a blueprint. We hope to obtain careful translations of FraCaS sentences into Portuguese. We plan to verify, via careful annotation work, that the (mostly logical) phenomena described in English remain the focus of the work and that the semantic phenomena 'work' in Portuguese the same way as they do in English. This is similar to the work in [1] for French and in the MultiFraCaS project[5]. A first draft translation can be found in the repo[6]. This preliminary translation was obtained via an online translation program, checked in the first instance by the authors, so that the sentences made sense in Portuguese. We have not verified the logic of the 339 problems, as yet. Our version was independently processed from the original one, instead of following McCarney's [7]. Recent work on inference has shown that human disagreements can persist [6], despite coordination between annotators. It will be interesting to see whether the semantic intuitions in FraCaS can be adopted wholesale in Portuguese, or whether they are invariant or not.

References

1. Amblard, M., Beysson, C., de Groote, P., Guillaume, B., Pogodolla, S.: A french version of the fracas test suite. In: LREC 2020-Language Resources and Evaluation Conference. p. 9 (2020)
2. Bowman, S.R., Angeli, G., Potts, C., Manning, C.D.: A large annotated corpus for learning natural language inference. In: Proc. of EMNLP. ACL, Lisbon, Portugal (Sep 2015), https://www.aclweb.org/anthology/D15-1075
3. Cooper, R., Crouch, D., van Eijck, J., Fox, C., van Genabith, J., Jaspars, J., Kamp, H., Milward, D., Pinkal, M., Poesio, M., Pullman, S.: Using the framework. the FraCaS Consortium (1996)
4. Fonseca, E.R., dos Santos, L.B., Criscuolo, M., Aluísio, S.M.: ASSIN: Avaliação de similaridade semântica e inferência textual. In: PROPOR. pp. 1–8 (2016)
5. Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., Zamparelli, R.: A SICK cure for the evaluation of compositional distributional semantic models. In: LREC 2014 (2014)
6. Pavlick, E., Kwiatkowski, T.: Inherent disagreements in human textual inferences. Transactions of the Association for Computational Linguistics **7**, 677–694 (2019)
7. Real, L., Fonseca, E., Oliveira, H.G.: The ASSIN 2 shared task: A quick overview. In: Quaresma, P., al (eds.) PROPOR 2020, Evora, Portugal. LNCS, vol. 12037, pp. 406–412. Springer (2020), https://doi.org/10.1007/978-3-030-41505-1_39
8. Real, L., Rodrigues, A., e Silva, A.V., Albiero, B., Guide, B., Thalenberg, B., Silva, C., Câmara, I.C.S., de Oliveira Lima, G., Souza, R., Stanojevic, M., de Paiva, V.: SICK-BR: a portuguese corpus for inference. In: PROPOR 2018 (2018)
9. Williams, A., Nangia, N., Bowman, S.R.: A broad-coverage challenge corpus for sentence understanding through inference. CoRR **abs/1704.05426** (2017), http://arxiv.org/abs/1704.05426

---

[5] https://gu-clasp.github.io/multifracas/

[6] https://github.com/vcvpaiva/fracas-BR-draft.csv

[7] https://www-nlp.stanford.edu/ wcmac/downloads/fracas.xml